IMPROVISATIONAL ARTIFICIAL INTELLIGENCE FOR EMBODIED CO-CREATIVITY

A Dissertation
Presented to
The Academic Faculty

By

Mikhail Jacob

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Interactive Computing

Georgia Institute of Technology

December 2019

IMPROVISATIONAL ARTIFICIAL INTELLIGENCE FOR EMBODIED CO-CREATIVITY

Approved by:

Dr. Brian Magerko, Advisor
School of Literature, Media, and Communication
Georgia Institute of Technology

Dr. Ashok Goel
School of Interactive Computing
Georgia Institute of Technology

Dr. Mark Riedl
School of Interactive Computing
Georgia Institute of Technology

Dr. Anne Sullivan
School of Literature, Media, and Communication
Georgia Institute of Technology

Dr. Mary Lou Maher
Department of Software and Information Systems
The University of North Carolina at Charlotte

Date Approved: July 29, 2019

The world is a slightly better place for having improvisation in it than it was before. There's something about it that says something positive about the human spirit, that a bunch of people can get together and by following a few simple traffic rules can create art and can entertain an audience and can thrill and exalt each other.

Del Close

The rules of improvisation apply beautifully to life. Never say no - you have to be interested to be interesting, and your job is to support your partners.

Scott Adsit

To Bridget, my deepest source of strength and hope.

ACKNOWLEDGMENTS

When I first came to Georgia Tech for my Masters in Computer Science in 2011, I was not sure what exactly I wanted to do immediately afterwards. I knew I wanted to work with intelligent systems or artificial intelligence in the software industry in some capacity but I was absolutely convinced of one thing – that I would never ever EVER want to pursue a doctoral degree. However, it seems that life and I had significant creative differences about that early decision. The incredible faculty and brilliant peers that I got to know in my graduate program at the time didn't just fan the flames of my intellectual curiosity but threw gasoline on it. Within a few short weeks I just knew that I too had to become a researcher. The dissertation that arose from that decision is a product of having attended an institution where it is completely ordinary to be surrounded by extraordinary people (both faculty and students) daily and where serendipitous explorations and curiosity-driven diversions are encouraged wholeheartedly. I am truly grateful for that enormous privilege.

There are far too many incredible people to thank for the direct and indirect roles that they played in getting me to the doctoral finish line with my dissertation in hand. I would like to start by thanking my advisor, Dr. Brian Magerko, who has guided and supported my research at the Expressive Machinery Lab for the past seven and a quarter years (starting well before my doctoral study itself did). His mentorship has been invaluable to me over the years in navigating not only the many unknowns of graduate school but also the twists and turns of life outside it. I am incredibly grateful for the amount of time, effort, funding, and motivation Brian has expended encouraging my research career.

My dissertation and the research leading up to it have been vastly improved through the insightful feedback from my supremely patient committee consisting of Dr. Ashok Goel, Dr. Mark Riedl, Dr. Anne Sullivan, and Dr. Mary Lou Maher. Their unwavering support and constructive criticism have challenged me, focused my research efforts, and elevated the quality of my dissertation numerous times. I would like to thank them for their guidance

and mentorship over the course of my doctoral journey.

The many years spent chugging along at the lab in the pursuit of this dissertation were made infinitely more enjoyable by the presence of many wonderful friends, colleagues, and collaborators — past and present; near and far. I was fortunate enough to get my fingers into as many research pies in the lab as I could and I would like to thank each of the phenomenal collaborators who I had the privilege to work with in the Expressive Machinery Lab (formerly the ADAM Lab). In particular, I would like to thank the following people for the absolute pleasure of working alongside them — The Computational Play Project/CoCoA (Justin Permar, Jonathan Streater, Eric Fruchter, Nicholas Davis), LuminAI/Viewpoints AI (Ivan Sysoev, Akshay Gupta, Allen Tsai, Margaret Hu, Sasi Viriyayuthakorn, Lauren Winston, Duri Long), TuneTable/BlockHead/GrooveMachine (Dr. Anna Xambó, Dr. Gerard Roma, Nikhil Bhanu, Anna Weisling, Ryan Rose, Dr. Astrid Bin), EarSketch Co-creative AI (Jason Smith, Dr. Jason Freeman, Dr. Tom McKlin), The Robot Improv Circus (Dor Hananel, Julia Vorpahl, Prabhav Chawla, Ziming He, Jason Lee, Michelle Ni, Lauren Douglas). In addition, I would also like to personally thank Jonathan Streater (for the constant friendship, wellspring of support, as well as the nightly debates on philosophy, cognition, and AI), Nicholas Davis (for the mentorship and for forcing me to broaden my mind to appreciate different intellectual perspectives), Dr. Astrid Bin (for the support and fantastic advice), and Duri Long (for the friendship, support, thoughtfulness, and just being there). A special acknowledgment is necessary for the wonderful doctoral students in and around my cohort who managed to share their support and excitement while simultaneously working through many of the same struggles (a huge thank you, Richard Rutledge, Michael Pettinati, Tesca Fitzgerald, Lara Martin, and Matthew Guzdial).

I consider myself immensely fortunate to have wonderful friends spread out across the world from all the different phases of my life. I would like to acknowledge the vital role that they played in helping keep my sanity intact throughout the doctoral process with their love, company, support, and humor. There are far too many lovely people in this list to

thank individually but I truly cherish each of those friendships, near or far.

It is difficult to express how grateful I am for the immense ocean of love and support that I have received from everyone in my large and extended family. Recently, I even gained a whole new extension to my family and I am incredibly grateful to my in laws, as well as to all of their extended family, for treating me like family from the very beginning. I would also like to acknowledge all the love and joy I received from my extended family spread out across the world, especially in India and the USA. They have been there for me through everything and I am supremely grateful to them. Finally, I would like to acknowledge the infinite reserve of love from my mother, father, sister, brother in law, and (most recently, my tiny) nephew that I have relied on to make it through the years. At times it has been unbearably difficult to be continents apart from them working through graduate school, a feeling that is intimately familiar to every international student and immigrant. However, their encouragement, patience, and unconditional love have brought me to this point in my life and made me who I am today.

There are no words that can do justice to the depth of my love and gratitude to my wife, Bridget. As I stumbled along the way and contemplated abandoning my doctorate several times, her strength and hope repeatedly helped carry me through to the conclusion of this grand doctoral journey. She has shared in all my highs and lows as a doctoral student — listening to me vent, supporting me through the numerous failures that are a part of research, offering advice when I needed it most, and celebrating all the joys of success with me. I can't begin to thank her for everything she has done and continues to do for me, including moving us across the world through sheer force of will. Finally, I would also like to thank my two kitties, Tilly and Boo, for providing endless hours of entertainment, adorable snuggles, and conditional love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Improvisation is an essential skill for co-creative agents to develop for successful performance despite resource constraints, time pressure, open-ended problems, and ill-defined goals. An important subset of improvisation that has diverse applications is embodied narrative improvisation, i.e., collaborative improvisation of narratives with other agents using the various modalities of its body situated within a virtual or physical environment. Unconstrained human-computer embodied narrative improvisation is a challenging problem since it requires the incorporation of many cognitive faculties including narrative intelligence (the ability to tell and understand stories), social cognition (reasoning about the goals, plans, desires of other beings), performance of linguistic/non-linguistic action (the physical ability to enact a set of actions), and commonsense reasoning (reasoning about how the world works at a naive level).

Unconstrained embodied narrative improvisation is too complex to address at present, as mentioned in the preceding paragraph. Therefore, this dissertation aims to explore the initial steps in a path toward improvisational agents that can eventually perform unconstrained embodied narrative improvisation with people. I focus on improvisation within an object-based gestural proto-narrative problem domain in this dissertation that falls under the set of problem domains that I collectively refer to as *movement improv* domains due to their focus on gestural and environmental interaction. My research addresses the *improvisational action selection problem*, which is a crucial challenge to creating improvisational agents in movement improv domains.

I study the *improvisational action selection problem* (the challenge of performing action selection from an open-ended action space with an ill-defined goal space in near real-time based on the agent's knowledge and the improvisational context, in order to avoid incoherent behavior, decision paralysis, and unexpressive responses) in this dissertation and how to address it within the *Robot Improv Circus* interactive virtual reality installation and the

CARNIVAL agent architecture. In this domain, *object-based gestural proto-narrative improvisation* takes place between non-expert human and virtual characters through the *Props game*. The CARNIVAL agent architecture uses *affordance-based action variant generation*, *improvisational response strategies*, and *computational evaluation of creativity* of perceived or generated actions to perform *creative arc negotiation* as a form of intrinsically motivated action selection in order to address the improvisational action selection problem. Creative arc negotiation is the process of selecting actions over time to best follow a given *creative arc*, i.e. a continuous target trajectory for generated responses to follow through an agent's *creative space* (the space of actions with different degrees of *novelty*, *surprise*, and *value*).

My dissertation has the thesis statement, "embodied agents that address the improvisational action selection problem using 'creative arc negotiation' increase perceptions of enjoyment, agent creativity, and coherence in both observers and participants while performing movement improv with non-experts." Through the evaluation performed in this dissertation, it was found that this thesis statement is valid to different extents as follows. It is valid to conclusively state that embodied agents addressing the improvisational action selection problem using creative arc negotiation can perform movement improv with non-experts so that perceptions of agent creativity and coherence increase for both participants and audience members, but that perceptions of enjoyment only increase conclusively for observers. More study and data is required to show a conclusive increase in perceptions of enjoyment for participants of the installation.

The contributions of my research in this dissertation are as follows.

- A model of affordance-based action variant generation for the parameterized generation of action variants based on a given objects physical attributes.
- A formalized set of improvisational reasoning strategies for guiding an agents action space search based on previous experience and the current improvisational context.

- Computational models for evaluating the creativity of perceived and generated action variants in terms of their novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value).

- A model of creative arc negotiation for improvisational action selection while performing movement improv with non-experts that increases both participant and observer perceptions of enjoyment, agent creativity, and coherence.

- A publicly disseminated and validated interactive installation where embodied agents can perform movement improv with non-experts.

CHAPTER 1
INTRODUCTION

Improvisation with human collaborators is an essential skill for intelligent agents to develop in order to act in realistically large-scale problem domains where cognitive, as well as physical resource limitations, severe time constraints, open-ended action spaces, and ill-defined goal spaces, are characteristic. For convenience, I have coined the term embodied narrative improvisation [1] to refer to an important set of creative domains within the space of improvisational domains that is at the intersection of embodied collaborative creativity (co-creativity) and narrative improvisation (in a broadly applicable sense). Successful human-computer embodied narrative improvisation would have valuable applications within diverse fields such as human-robot (or human-agent) interaction, immersive scenario-based training, expressive arts or play therapies, virtual reality (VR) games or entertainment, and using performing arts to encourage broader participation in STEM (science, technology, engineering, and mathematics). Embodied narrative improvisation, as I defined it in [1], involves an agent co-constructing and enacting narratives with other agents using the various interaction modalities, constraints, and affordances of its body situated within a virtual or physical environment. For example, within a VR game that enables players to interact with the virtual world and each other through naturalistic embodied interaction, non-player characters (NPCs) could co-construct the game's narrative with players through their embodied actions instead of being restricted to following scripted sequences of canned animations and pre-recorded behaviors.

The problem of human-computer embodied narrative improvisation is too complex and challenging to address in its unconstrained form because it requires agents to possess many complex reasoning capabilities such as narrative intelligence (the ability to tell and understand narratives, see section 1.2), social cognition (the ability to reason about other agents'

mental states and how that interacts with one's mental state), performance of linguistic and non-linguistic action, as well as many other reasoning faculties, interaction capabilities, knowledge, and experience. Therefore, this dissertation aims to form the first exploratory steps towards someday creating improvisational agents that can perform unconstrained embodied narrative improvisation with people. In order to achieve this, I have simplified the scope of my research in several ways. Firstly, due to the relative abundance of research in speech-based and textual narrative domains, I focus on studying improvisation within a problem domain where the primary interaction modalities are object-based and gestural interaction. I refer to this set of simplified problem domains as *movement improv* domains (see section 1.2) throughout this dissertation in order to emphasize that they primarily involve full-body gesture and object-based interactions. Secondly, I restrict the scope of the research in this dissertation to addressing the *improvisational action selection problem* (see section 1.2.1), which is a key challenge for creating improvisational agents for movement improv that prior research in improvised dance [2] and pretend play [3] had highlighted. Finally, as a tangible step closer to unconstrained embodied narrative compared to my prior work in human-computer improvised dance [2], in this dissertation I choose to study the improvisational action selection problem within a problem domain that represents an increase both in the complexity of the improvised embodied interactions and in the degree of semantic structure required for successful improvisation.

The remainder of this chapter starts by introducing the terms and concepts used in this research before describing the improvisational action selection problem. The chapter then discusses the domain chosen to study the improvisational action selection problem as well as the techniques used to create improvisational agents within that domain. The chapter then presents my thesis statement and the research questions guiding the formal evaluation of the claims in my thesis statement. This chapter finally concludes by detailing the contributions of this research.

1.1 Creativity and Improvisation

This section formally introduces the concepts of creativity, co-creativity, and improvisation as they are used in this dissertation. The definitions for these terms are required to contextualize the main problem addressed through this research (see section 1.2.1) and the computational techniques used to address it (see section 1.4). More detail about these concepts and relevant related research can also be seen in sections 2.1 and 2.4.

Formal research into the phenomenon of creativity from both humanistic and computational perspectives have has led to a scientific understanding of human creativity including recommendations for improving creativity [4], better supporting creative practice [5], and scaling up creative impact [6]. Creativity research has also resulted in several definitions of creativity as a phenomenon. Newell, Shaw, and Simon's creative problem solving [7] referred to elements of creativity as novelty, value, rejection of previous assumptions, persistence towards a goal, and the development of a problem specification itself. Boden's influential model [8] of creativity defined creativity as "the ability to come up with ideas or artifacts that are new, surprising and valuable" (under various senses of the terms 'novelty', 'surprise', and 'value'). Colton's creativity tripod [9] argued for creativity involving skill, imagination, and an appreciation of a chosen creative medium. Colton, Charnley, and Pease's later FACE model [10], on the other hand, defined conceptual creativity in terms of creative concept invention, expression of the concept as an artifact, aesthetic evaluation of the artifact, and the framing of the artifact to an audience. Finally, Jordanous' Standardized Procedure for Evaluating Creative Systems (SPECS) methodology [11] described a three-part methodology for evaluating creativity that provided fourteen common criteria that were commonly associated with defining creativity ranging from originality to domain competence and value. More detail about how these definitions contribute to computational models for evaluating creativity can be found in section 2.4. For this dissertation, Boden's [8] product-based definition of creativity (mentioned above) is adapted and operationalized

for the current context in the following definitions along with component terms such as novelty, surprise, unexpectedness, value, and quality.

Definition 1.1.1 (Novelty) *The aggregated difference between a percept or artifact and other comparable experiences or artifacts that an agent has already experienced.*

Definition 1.1.2 (Unexpectedness) *The degree that an experience or artifact deviates from the agent's expectation for that experience or artifact.*

Definition 1.1.3 (Surprise) *An affective reaction to an experience or artifact caused by the violation of confidently held expectations about that experience or artifact proportional to the degree of experienced unexpectedness.*

Definition 1.1.4 (Quality) *The standard of an experience or artifact in comparison to other comparable experiences according to specific, predetermined criteria used for assessment.*

Definition 1.1.5 (Value) *The usefulness and quality of an experience (or artifact) to the creator(s), consumer(s), embedding society or culture, and the contexts for the creation, consumption, and gatekeeping of that experience (or artifact).*

Definition 1.1.6 (Creativity) *The creativity of an artifact is the weighted aggregation of its novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value) as experienced by an evaluating agent in relation to its past experiences, current expectations, and quality criteria.*

The computational creativity research community has traditionally focused on the autonomous generation of creative artifacts [12], in addition to their contributions towards modeling creativity. Recently, there has been a rising interest in *collaborative creativity* or *co-creativity* between human and machine [13]. Co-creativity is defined in this work with a process-based perspective as follows (see more about different kinds of perspectives in section 2.4).

Definition 1.1.7 (Co-creativity) *The process by which two or more agents (human or computer) collaborate within a creative domain in a variety of possible configurations and collaborative roles to generate a creative artifact together.*

Co-creative agents can assist, augment, direct, or otherwise relate to human creativity by taking on a variety of roles in the creative collaboration. These roles can include creativity support tool [13], creative task worker [13], creative assistant [13], inspirational source [13], nanny [14], coach [14], pen-pal [14], colleague [14], critic [15], task provider [16], or instructor [16], task leader [16], task follower [16]. Newer co-creative agents can also transition between different co-creative roles in the co-creative process depending on the context over time (e.g. transitions between leader-follower roles in [17]).

Improvisation is a term that is used to refer to a broad spectrum of creative domains that involve the production of creative outputs 'in the moment' to varying degrees. Berliner [18] describes improvisation (in jazz) as "reworking pre-composed material and design in relation to unanticipated ideas conceived, shaped, and transformed under the special conditions of performance, thereby adding unique features to every creation." Pressing [19] describes how the improvisational process proceeds with respect to a "formal schema or guiding image" called the *referent* that serves as inspiration or constraining criteria throughout the improvisational performance. According to Sawyer [20] the degree of guiding structure in an improvised performance can vary drastically based on the performance domain, ranging in complexity from improvisation that is "as basic as a performer's elaboration or variation of an existing framework a song, ritual prayer, or traditional story," to those improvisational performances where "the performers start without any advance framework and create the entire work on stage." However, across these two extremes of improvisational complexity, the improvisational process highlights the impressive ability of a creative agent (human or computer) to fluidly generate creative responses in near real-time within open-ended and ill-defined problem domains.

Improvisational creativity necessarily operates using constrained cognitive and physical

resources, under severe time constraints, over a potentially unbounded action space (i.e., an open-ended action space), and without a single (or small set of) well-defined goal(s) to pursue at any given time (i.e., it has an ill-defined goal space). The improvisational domains referred to or addressed in this dissertation are characteristically collaborative (i.e. involving more than one improvising agent), open-ended (i.e. having a potentially unbounded action space from which to select creative responses in near real-time), ill-defined (i.e. lacking a clear set of well-specified goals to follow or objective functions to optimize in order to select responses in near real-time), and performative (i.e. restricted to creative domains where some set of agents is performing for an audience in near real-time). Therefore, improvisation is defined within this work as follows.

Definition 1.1.8 (Improvisation) *Improvisation is the process of collaboratively producing creative outputs in near real-time within open-ended, ill-defined creative performance domains.*

Improvisation has long been studied in human creative practice and creative process [19, 21, 22, 23]. This has most often taken the form of observational studies of human improvisers. Computational models of improvisational creativity have also been explored for a small number of domains, including music [24], visual art [25], pretend play [3], theater [26], and emergency response management [21].

Research in computational creativity, co-creativity, and improvisation have largely ignored domains of embodied creativity such as dance, theater, mime, and pretend play [27] (though notable exceptions exist). In addition to being artistically and creatively important fields, it is a particularly opportune time to study embodied co-creativity with the easy availability of cheap, high quality body sensing technology enabling reliable embodied interaction and the mass-market adoption of VR technology highlighting the transformative potential for embodied co-creativity in a wide range of extended reality (XR) applications. This dissertation thus focuses on research into human-computer improvisation within domains highlighting embodied co-creativity.

1.2 Embodied Narrative Improvisation and Movement Improv

The human affinity for understanding the world and communicating ideas through narrative is referred to as narrative intelligence (see [28, 29, 30]). As a central human faculty, narrative has been defined in many ways according to various schools of thought over time. Abbott [31] presents a minimal definition of narrative as, "the representation of an event or a series of events." Graesser, Hauft-Smith, Cohen, and Pyles [32] defines narrative as prose that "delineates actions and events which causally unfold in time, e.g., stories and tales." Graesser, Singer, and Trabasso [33] add that narratives "involve people performing actions in pursuit of goals, the occurrence of obstacles to goals, and emotional reactions to events." Lakoff and Johnson [34] detail the features that narratives commonly possess in their 'Life Is A Story' metaphor, including participants (characters), parts (settings, episodes, states etc.), stages (temporal sections of the story), linear sequences (temporal and/or causal relations between successive episodes and states), causation (causal relations between episodes and states), and purpose (goals and plans). Following from these perspectives, narrative is defined in this research using Prince's [35] definition as follows.

Definition 1.2.1 (Narrative) *The representation of at least two real or fictive events in a time sequence, neither of which presupposes or entails the other.*

Embodied narrative refers to the physical (or virtual) enactment of narrative grounded in an agent's embodied experience, constructed using its body within the physical (or virtual) environment in which it is situated [36]. Some examples of embodied narrative include reenacting a favorite movie, acting in the theater, playing certain virtual reality games, and performing a dance interpreting a classic story. Disembodied narrative, on the other hand, includes a text translation of the Epic of Gilgamesh in English or an autobiographical blog post on the world wide web. Embodied narrative is defined in this work as follows.

Definition 1.2.2 (Embodied Narrative) *A narrative that a physically (or virtually) embodied agent constructs using the interaction modalities, constraints, and affordances of*

its physical (or virtual) body as well as interactions with its physical (or virtual) environment.

Embodied narrative improvisation encompasses a challenging class of improvisational domains, such as long-form improvisational theater, live-action role-playing (LARP) games, collaborative pretend play, and certain forms of improvisational dance. These activities involve embodied narrative co-construction predominantly in real-time based on the current improvisational context and the creative offers (opportunities for progressing the narrative) being passed back and forth between improvisers. The complexity of the improvisational task in a particular domain may be constrained by differing levels of structure arising from the 'rules' or conventions of the domain. Embodied narrative improvisation is defined in this work as follows.

Definition 1.2.3 (Embodied Narrative Improvisation) *Embodied narrative co-construction among multiple participants in near real-time by performing actions to advance the narrative from an open-ended narrative action space and an ill-defined goal space.*

The opportunity for embodied narrative improvisation to make a lasting impact has been highlighted in recent years by the explosion of immersive VR for entertainment and industrial applications. VR-focused applications, including VR games, productivity tools, and training experiences, are a fast-growing segment of the digital entertainment [37], artistic practice [38], as well as training and simulation industries [39, 40]. The embodied interaction in VR, such as walking and manipulating the virtual world using the body, strongly supports a user's ability to perform embodied narrative improvisation and adds to a user's immersion [41]. Users can physically mime flipping burgers [42] or scaling Mt. Everest [43] to do so in-game. Naturalistic embodied interaction in VR, thus enforces the direct correspondence between the movements performed by their body and the character's actions in the virtual world increasing user presence [41].

There is currently a jarring imbalance in the agency available to (and responsibility placed on) the human collaborator in VR games and experiences at present. For example, in the game Job Simulator [42], the user has extraordinary physical control of the world using their body to flip burgers, stamp sales reports, and even play Minesweeper on a computer. However, the other NPCs in the game are all floating computer heads with disembodied hands that can float in predefined points and do the same actions the same way each time. This is because a non-player character (NPC) in VR experiences is still limited to using sets of pre-recorded animations (albeit cleverly blended in predetermined ways) to move or act. In Job Simulator, the distinct lack of variation can be excused since the other NPCs are presented as floating CRT monitors with disembodied hands, and their robotic behavior is aesthetically appropriate. However, any VR game or experience that aims to portray a behaviorally realistic (or behaviorally believable) human (or humanoid) NPC needs to overcome this lack of character expressiveness and fully support open-ended embodied interaction [44].

Open-world sandbox games in VR like [45, 46] or expressive interactive experiences in embodied environments like [47, 48, 49] are usually solitary explorations of human creativity and expression or rely on human players to provide a sense of social gameplay. This is because developing NPC AI for co-creativity and expression is vastly more complex than doing so for task-oriented games with a fixed set of rules and clear rewards for following them. The lack of clearly defined goals to perform at each point in these open-ended games makes the experience of playing them more akin to an improvisational narrative than traditional narratives evolving from a fixed set of possible actions and ways that they can be performed. Therefore, NPCs that could improvise in embodied environments could form collaborative companions to players in open-world sandbox games and expressive interactive experiences in embodied environments.

The grand challenge of creating a computational agent that can perform unconstrained embodied narrative improvisation with people in a real-world creative domain would re-

quire the agent to possess models of narrative intelligence, social cognition, common sense reasoning, meta-communication, and many other cognitive faculties in addition to vast amounts of knowledge about the creative domain and experience within it. Moving along the path towards unconstrained human-agent embodied narrative improvisation, the research in this dissertation continues from prior work on gestural proto-narrative (a sequence of temporally and aesthetically related actions executed by a set of actors [50]) and dance improvisation in the LuminAI installation [50, 2, 16] to investigate object-based proto-narrative improvisational theater. Object-based proto-narrative improvisational theater was chosen because of its emphasis on embodiment and environmental interaction. Improvisation within this domain would also demonstrate success despite an increase both in the complexity of the interaction modalities used for improvisation and in the degree of narrative (as well as semantic) structure required for the domain, compared to prior work in the LuminAI installation. Additionally, while unconstrained embodied narrative improvisation could be performed through any of the body's interaction modalities including speech, gesture, non-verbal communication, environmental interactions, this research is restricted to embodied improvisation using gesture and object-based interactions within virtual environments in order to make the problem tractable (as in my prior work). These restricted domains of embodied improvisation that are closely related to embodied narrative improvisation and include domains from prior work as well as the main problem domain for research in this dissertation are collectively referred to as *movement improv*.

Definition 1.2.4 (Movement Improv) *The set of embodied improvisational domains that are closely related to embodied narrative improvisation but are restricted to focus on full-body gestural interaction between fellow improvisers as well as object-based interactions within the environment they are situated in while requiring varying levels of improvisational complexity in terms of narrative and semantic structure.*

Movement improv forms a simplified set of embodied improvisational domains compared to unconstrained embodied narrative improvisation. However, it encompasses a large

number of creative domains and practices from the real-world such as improvisational dance, prop-based improv theatre games, collaborative VR sandbox games, or pretend play with toys. Human-agent movement improv is difficult for improvisational agents to perform due to several inherent challenges. Prior work in human-agent movement improv within the domains of pretend play [51] studied the problem of narrative improvisation with non-experts. More recent work in gestural proto-narrative and improvisational dance within the LuminAI installation [50, 2, 16] explored how to address the knowledge-authoring bottleneck (the difficulty of acquiring expert knowledge followed by its subsequent representation and storage to enable efficient future utilization) that restricted the pretend play research to severely limited problem domains for improvisation. My prior research within the LuminAI installation highlighted that a critical challenge for embodied improvisational agents to perform human-agent movement improv is the improvisational action selection problem.

1.2.1 The Improvisational Action Selection Problem

Embodied agents that are performing movement improv with people are required to produce a response in near real-time based on the current context of the unfolding improvised performance. They face the challenge of selecting their response from an open-ended action space in the presence of an ill-defined goal space to guide their action selection. In order to work within the severe temporal constraints for action selection within improvisational domains, the agent could attempt to use stochastic or shallow reasoning. However, this can easily result in generated agent behavior that is perceived as incoherent in the long term [2, 16]. On the other hand, attempting to perform deep and complex reasoning in order to select an action can easily result in violating the temporal constraints of the domain, leading to perceived decision paralysis from the agent. In both cases, the absence of well-defined goals to follow or objective functions to optimize for action selection also adds to the risk of the selected actions not being meaningfully different or expressive

enough at different points in the improvised performance or across different performances. Thus the agent is required to perform improvisational reasoning in order to select actions while avoiding incoherent behavior, decision paralysis, and unexpressive responses. The improvisational action selection problem is then defined as follows in this research.

Definition 1.2.5 (The Improvisational Action Selection Problem) *The challenge of performing action selection as an improvisational agent in near real-time from an open-ended action space with an ill-defined goal space based on previous experience and the current improvisational context in order to avoid incoherent behavior, decision paralysis, and unexpressive responses.*

The severity of the improvisational action selection problem is directly dependent on the complexity of the improvisational task. Pressing [19] refers to this in musical improvisation as "a continuum of possibilities between the extreme hypothetical limits of 'pure' improvisation and 'pure' composition." He also states that for human improvisers the two theoretical extremes are "never obtained in live performance because no improviser (even in 'free' improvisation) can avoid the use of previously learned material, and no re-creative performer can avoid small variations specific to each occasion." The complexity of improvisation directly affects the severity of the improvisational action selection problem with 'pure' improvisation being the most challenging to address. For a severely reductive example, the use of jazz standards reduces the improvisational complexity to the task of improvising melodic variations based on these widely-known songs. On the other hand, the task of improvising a long-form improvisational theater narrative consists of a much more fundamental negotiation of the conventions for a performance in parallel with the improvisation of content within those negotiated conventions. The latter is, therefore, a more complex improvisational task and faces a more severe form of the improvisational action selection problem.

The complexity of the improvisational task and the severity of the improvisational action selection problem vary inversely with the degree of formal structure or improvisational

constraints present on that task. While not a perfect mapping, the degree of structure in Pressing's [19] concept of referent (an "underlying formal scheme or guiding image specific to a given piece" that is "used by the improviser to facilitate the generation and editing of improvised behaviour") is useful as a proxy to understand the degree of structure present in a given improvisational task. The more rigid the productive mapping between referent and generated or integrated content, the more tightly constrained and more structured the improvisation. Pressing [19] also states conversely, that "if no referent is present, or if it is devised in real-time," the result is "/,'free' or 'absolute' improvisation." This is a much rarer improvisational form than the previous "referent-guided, or 'relative' improvisation" and faces a much more severe improvisational action selection problem. For example, comparing modern improv theater with the Commedia Dell'arte [52], there are clearer constraints imposed on improvisation within Commedia Dell'arte due to the stereotypical characters and high-level plot outlines than on modern improvisational actors, leading to a greater severity of the improvisational action selection problem in modern improv theater as a domain than the Commedia Dell'arte.

Various computational techniques have been used in the past for problem spaces where it is not desirable (or even possible necessarily) to define/enumerate a set of goals for an agent to follow. Some of these techniques include evolutionary algorithms and reinforcement learning (along with its variations). For movement improv, however, by definition, it is not possible to formalize the entire problem into a set of objective function(s) for evolutionary approaches to optimize. Similarly, reinforcement learning (RL) [53] is not a feasible solution either due to the lack of a well-specified reward function for movement improv. A number of inverse RL [54] or imitation learning [55] variants from the RL research space could potentially be used in this situation since they either learn a reward function to optimize or directly learn a policy imitatively respectively from observing humans complete a task. However, due to the vast size of the action space for performing movement improv and the sample inefficiency of these approaches, they cannot practically

be used at the moment.

1.3 Problem Domain

The research in this dissertation continues my investigation of improvisational agents for embodied co-creativity in movement improv from prior work that focused on how agents can improvise gestural proto-narrative or dance while attempting to address the knowledge-authoring bottleneck involved [50, 2, 16]. The problem domain described in this dissertation for studying the improvisational action selection problem shows an increase in improvisational complexity from my prior work in order to form a clear progression towards unconstrained embodied narrative improvisation in the future.

1.3.1 Object-based Gestural Proto-narrative and The Props Game

Object-based gestural proto-narrative (a sequence of temporally and aesthetically related actions executed by a set of actors [50] using the interaction modalities of gestural and object-based interaction) within a virtual environment was explored as the movement improv domain for this dissertation continuing from prior work in gestural proto-narrative and improvisational dance [50, 2, 16]. This particular problem domain was chosen as an advancement over that prior work in terms of both the degree of narrative (or action semantics) involved and the interaction modalities for the agent to improvise with people. The 'rules' of the domain were also structured enough to allow for the exploration of the improvisational action selection problem.

The specific form of object interaction-based proto-narrative that was chosen for this research was the Props Game from short-form improv theatre. The Props Game involves improvised interactions between two or more participants using unfamiliar, ambiguous props to perform recognizable comedic actions pretending the prop to be a familiar real-world or fictional object. Therefore, in this research, the performance was taking place between an embodied virtual agent and a human improviser using ambiguous props that

were potentially unfamiliar to the agent. Beyond improv theater, this problem is useful for embodied agents in general since it is the first step towards allowing them to gain new knowledge about unfamiliar objects through interaction. For example, this could include an agent learning to use unfamiliar objects in unfamiliar scenarios according to familiar human norms/customs or using unfamiliar objects for a specific task, such as improvising a digging tool for disaster recovery.

1.3.2 Movement Improv With Non-experts

Improvisation with non-experts was explicitly a design consideration in this research for multiple reasons. Firstly, previous research [56] had indicated the reduced dissemination impact of the improvisational experiences when restricting the experiences and activities that were involved to a target population of experts in a niche domain. Secondly, expert users tended to want to exert more control over the improvisational performance and in co-creative interactions than non-experts did [13]. Thirdly, it was intended that designing installations in public spaces for non-experts would democratize access to the installation and encourage more diversity in the knowledge that was learned and eventually entered the installation. Finally, it was also decided that non-expert data would be used to train the agents in the installation in order to reduce participants' social embarrassment about improvising in public next to an intimidatingly expert improviser and lower the barrier for entry to participate in the installation.

1.4 Improvisational Agents For Performing Movement Improv

Improvisational agents that are required to perform movement improv with non-experts need to address the improvisational action selection problem (see section 1.2.1). Perhaps improvisational agents can take inspiration (and generalize) from the different aesthetic trajectories that are found to give guiding structure to various artistic and creative domains such as dramatic arcs in narrative, arcs of rising or falling tension in music, and visual lines

or movement in visual art in order to address the improvisational action selection problem and perform movement improv with non-experts. I hypothesize that **improvisational agents performing movement improv with non-experts using a form of intrinsically motivated action selection called *creative arc negotiation* successfully address the improvisational action selection problem.** Creative arc negotiation and related terms are defined in this dissertation as follows.

Definition 1.4.1 (Creative Space) *The multi-dimensional space of novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value) within which perceived or generated action variants can be localized.*

Definition 1.4.2 (Creative Arc) *The desired temporal progression or target trajectory for an agent's selected actions within a creative space over the course of an improvised performance.*

Definition 1.4.3 (Creative Arc Negotiation) *Interruptible, temporally constrained, search-based action selection to best follow a given creative arc through the agent's creative space considering the current improvisational context and the agent's previous experience.*

Addressing the improvisational action selection problem entails by definition that the agent can select actions in near real-time from an open-ended action space and an improvisational domain with an ill-defined goal space. This also implies that the improvisational agent can successfully avoid decision paralysis, incoherent behavior, and unexpressive responses if it is able to follow the guiding structure of the given creative arc over the course of the improvised performance. Therefore, I hypothesize that **successfully addressing the improvisational action selection problem using creative arc negotiation will increase both participant and audience perceptions of enjoyment, agent creativity, and coherence over the course of the improvised performance.**

An improvisational agent performs creative arc negotiation by strategically searching an action space during its turn and evaluating candidate action variants that are generated

during the search to find the closest match to the next target point on the given creative arc within the temporal constraints of its turn. Through the research in this dissertation, I explore how improvisational agents can operationalize the creative arc negotiation process using the following components.

- A parameterizable action variant generator from the agent's action space.
- A set of improvisational reasoning strategies for guiding the agent's action space search based on previous experience as well as the current improvisational context.
- A set of computational models for evaluating the creativity of perceived or generated action variants in terms of their novelty, unexpectedness, and quality.

An improvisational agent that aims to perform movement improv with non-experts using creative arc negotiation needs to be able to generate action variants from the agent's action space based on a given set of parameters. In the object-based gestural proto-narrative domain described in this dissertation, the physical attributes of objects that are provided to the agent to use for improvising actions form a useful set of parameters to conditionally constrain the action variant generation in addition to other search parameters. A computational model that learns a mapping between the physical attributes of objects and the set of possible actions with those objects enables the agent to generate action variants that can serve as candidate actions for the agent to use as its responses. Such a model implements an acquired relation between the agent's learned action space, the physical object attributes that the agent has experienced, and the embodied capabilities of the agent. Therefore, it forms a model of affordance-based action variant generation according to Şahin, Çakmak, Doğar, Uğur, and Üçoluk [57]'s definition of affordance (see sections 2.3.1 and 3.3.4 for more detail). This work explores how **affordance-based action variant generation enables the agent to perform parameterized action variant generation from a learned action space as a part of creative arc negotiation.**

Creative arc negotiation requires the agent to search its creative space to find the closest candidate response to the next target point on the given creative arc in near real-time. Human improvisers have been known to use various reasoning strategies for generating responses in near real-time based on their previous experience and the current improvisational context across various forms of improvisation [19, 58]. Computational formalizations of these improvisational reasoning strategies have been used in prior work to enable the agent to respond to humans with potentially valid actions even when it has not learned the constraints or 'rules' of the domain [50]. This research examines how **improvisational reasoning strategies formalized from human improvisers and extended from prior work enable the agent to search its action space based on its previous experience and the current improvisational context while performing creative arc negotiation.**

Creative arc negotiation requires the improvisational agent to computationally evaluate the creativity of action variants that it generates as possible responses to its partner as well as the human improviser's actions that it perceives. The agent evaluates these actions in terms of their novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value). This dissertation studies how **computational models of novelty, unexpectedness, and quality enable the agent to evaluate the creativity of perceived and generated actions in near real-time for performing creative arc negotiation.**

1.5 Thesis Statement

This dissertation synthesizes the hypotheses described above and presents an investigation of the following thesis statement.

> Embodied agents that address the improvisational action selection problem using 'creative arc negotiation' increase perceptions of enjoyment, agent creativity, and coherence in both observers and participants while performing movement improv with non-experts.

1.6 Research Questions

The avenues of inquiry presented by the thesis statement above were pursued through the following research questions (RQ).

RQ1 How can an agent perform parameterized action variant generation from a learned action space based on the physical attributes of a given object?

RQ2 How can an agent improvisationally search its action space based on previous experience and the current improvisational context?

RQ3 How can an improvisational agent computationally evaluate the creativity of perceived or generated actions in near real-time in terms of their novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value)?

RQ4 How can an embodied agent select actions to negotiate a given creative arc in order to address the improvisational action selection problem while performing movement improv with non-experts?

RQ5 How does addressing the improvisational action selection problem while performing movement improv with non-experts affect both observer and participant perceptions of enjoyment, agent creativity, and coherence?

The research in this dissertation investigating the guiding research questions (RQ) is outlined below as a set of objectives (O) for exploring each research question, methods (M) used to achieve the objectives and measurable outcomes (MO) from the research in the following list.

RQ1 How can an agent perform parameterized action variant generation from a learned action space based on the physical attributes of a given object?

O1.1 To create a computational model of parameterized action variant generation from a learned action space, conditioned on the physical attributes of objects (affordance-based action variant generation).

* Methods

M1.1.1 Explore the use of conditional variational autoencoders and latent space sampling for affordance-based action variant generation.

M1.1.2 Validate the model of affordance-based action variant generation.

* Measurable Outcomes

MO1.1.1 A validated model of affordance-based action variant generation from a learned action space.

RQ2 How can an agent improvisationally search its action space based on previous experience and the current improvisational context?

O2.1 To formalize procedural reasoning strategies adapted from human improvisation practices across domains in order to guide the agents action selection in open-ended action spaces with ill-defined goals using previous experience and the current improvisational context.

* Methods

M2.1.1 Explore the use of procedural strategies for latent space search within a conditional variational autoencoder model to formalize improvisational response strategies from human improvisers and search the agent's learned action space using previous experience and the current improvisational context.

M2.1.2 Evaluate the agent can perform improvisational action selection in open-ended action spaces with ill-defined goals using improvisational response strategies.

* Measurable Outcomes

MO2.1.1 Validated formalization of improvisational response strategies that heuristically guide the agents action selection in open-ended action spaces with ill-defined goals using previous experience and the current improvisational context.

RQ3 How can an improvisational agent computationally evaluate the creativity of perceived or generated actions in near real-time in terms of their novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value)?

O3.1 To create a computational model for evaluating the novelty, unexpectedness, and quality of perceived human actions and generated agent actions.

* Methods

M3.1.1 Explore the use of content-based mean distance to evaluate the gestural and semantic novelty of perceived or generated actions.

M3.1.2 Explore the use of Bayesian Surprise [59] and distance from expected outcome [60] to evaluate the gestural and semantic unexpectedness of perceived or generated actions given the physical attributes of the object being used to enact them.

M3.1.3 Explore the use of heuristic functions to evaluate the quality of perceived or generated actions in terms of their smoothness and recognizability.

M3.1.4 Evaluate whether the models for evaluating the novelty, unexpectedness, and quality of perceived and generated actions match human perceptions of these qualities.

* Measurable Outcomes

MO3.1.1 Validated computational model for evaluation of novelty for human and agent actions.

MO3.1.2 Validated computational model for evaluation of unexpectedness for human and agent actions.

MO3.1.3 A validated computational model for evaluation of quality in terms of the smoothness and recognizability of the gesture

RQ4 How can an embodied agent select actions to negotiate a given creative arc in order to address the improvisational action selection problem while performing movement improv with non-experts?

O4.1 To create an embodied agent architecture that enables an agent to negotiate a given creative arc while performing movement improv with non-experts.

* Methods

M4.1.1 Explore the use of parameterized action variant generation, formalization of improvisational response strategies, and creativity evaluation models to enable an agent to negotiate a given creative arc while performing movement improv with non-experts.

M4.1.2 Evaluate whether the embodied agent architecture enables an agent to negotiate a given creative arc while performing movement improv.

* Measurable Outcomes

MO4.1.1 A validated embodied agent architecture that enables an agent to negotiate a given creative arc while performing movement improv with non-experts.

RQ5 How does addressing the improvisational action selection problem while performing movement improv with non-experts affect both observer and participant perceptions of enjoyment, agent creativity, and coherence?

O5.1 To evaluate how an embodied agent that can negotiate a creative arc while performing movement improv with non-experts affects user and observer percep-

tions of enjoyment, agent creativity, and coherence.

* Methods

M5.1.1 Evaluate how participant and observer perceptions of enjoyment, agent creativity, and coherence are affected when improvising with a creative arc negotiating embodied agent within an interactive installation for performing movement improv with non-experts.

* Measurable Outcomes

MO5.1.1 Validated results on how an embodied agent that can negotiate a creative arc while performing movement improv with non-experts affects participant and observer perceptions of enjoyment, agent creativity, and coherence.

1.7 Contributions

The contributions of the research in this dissertation are as follows.

- A model of affordance-based action variant generation for conditionally searching the agents learned action space based on a given objects physical attributes.
- A formalized set of improvisational reasoning strategies for guiding an agents action space search based on previous experience and the current improvisational context.
- Computational models for evaluating the creativity of perceived and generated action variants in terms of their novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value).
- A model of creative arc negotiation for improvisational action selection while performing movement improv with non-experts that increases participant and observer perceptions of enjoyment, agent creativity, and coherence.

- A publicly disseminated and validated interactive installation where embodied agents can perform movement improv with non-experts.

The remainder of this dissertation continues by detailing related work that contextualizes the research conducted and the claims made in this dissertation. The dissertation then describes the improvisational action selection problem and how it is studied within the Props game domain through the Robot Improv Circus VR installation and the CARNIVAL agent architecture. The chapter continues to discuss the problem studied, the framework for addressing it in relation to my thesis statement, the technical approach taken to address the problem, various evaluation experiments for understanding the degree to which the solution addressed the claims in my thesis statement, and a further discussion to highlight additional insights from the system building and evaluation process. Finally, the dissertation concludes with a chapter that reiterates the contributions of the research in this dissertation and provides a quick sketch of future directions for this work as it relates to the problem of unconstrained embodied narrative improvisation.

CHAPTER 2

RELATED WORK

2.1 Creativity Research and Computational Creativity

Formal studies of creativity have focused on different aspects of the phenomenon. Some researchers have studied creativity from a domain-independent perspective considering the psychometrics of creativity [61], characterizations of creative personalities [62], analyses of creative environments [63], case studies of creative process [64], and experimental enumeration of the cognitive processes involved in creative cognition [65]. Others have studied creativity observing the processes that practitioners in specific creative domains actually follow to produce creative artifacts. These include studies of artistic or expressive domains like music [66], visual art [67], and storytelling [68] as well as other creative domains such as design [69], insight problem solving [70], and scientific invention [71].

The fields of artificial intelligence and machine learning have, more recently, given creativity researchers the tools and techniques to computationally model creativity. These have included computational systems that either attempted to emulate human creative cognition or to use uniquely computational processes for performing creative tasks. Some examples include MEXICA [72] system that used a cognitive model of human composition (originally from creative writing) called engagement-reflection (ER) to generate stories, the COLIBRI [73] system that used a case-based reasoning (CBR) [74] approach to generate poetry, and the Painting Fool [75] artificial visual artist system that used several models of visual art and visual creativity to generate paintings.

My research into improvisational agents for embodied co-creativity is situated within this field of computational creativity. Computational creativity research involves the "philosophy, science and engineering of computational systems which, by taking on particular

responsibilities, exhibit behaviors that unbiased observers would deem to be creative" [76]. Computationally creative systems include the Angelina system for automatically generating video games [77], The Painting Fool system for generating visual art [75], IDyOM for music generation [78], and COLIBRI for poetry generation [73]. This traditional definition has been expanded over the years, bringing related areas of research into the fold. A relatively recent area of interest within the computational creativity community is in computational systems that perform co-creation and co-creativity alongside human collaborators.

2.1.1 Computationally Co-creative Systems

Co-creativity (or collaborative creativity) refers to creative processes where there is active participation from two or more collaborators at a high-level [79]. Additional criteria can be applied to this definition such as the need for synchronous participation and collaborative emergence [80], the relative balance of creative agency and responsibility [2], or various creative roles for the collaborators in the co-creative process [14, 16, 17]. Some authors have also identified co-creation as a more general process that reduces the creative responsibility of the computational agent involved in the collaboration [13]. Some examples of co-creative or co-creational agents in the literature include collaborative sketching agents [25, 81], collaborative game design agents [82, 83, 84, 85], creative writing or storytelling [86, 87, 88, 89, 90], and dance or choreography [91, 92, 93, 94, 95]. The vast majority of existing co-creative agents in the literature are disembodied agents that users interact with through software user interfaces. In contrast, this dissertation focuses on understanding how to build co-creative agents specifically for embodied improvisational domains such as gestural proto-narrative, dance, and object-based gestural proto-narrative improvisation. Additionally, the research presented in this dissertation contributes to co-creative agents that possess equal creative agency and responsibility during the co-creative process.

2.1.2 Improvisational Systems

Improvisational agents demonstrate near real-time co-creativity or co-creation in open-ended, ill-defined domains. A limit set of previous improvisational systems can be found in the domains of musical improvisation [96, 97, 98, 99], emergency response management [100], improvisational theater [101, 56, 26, 102, 103], collaborative visual art [104], dance [105], and improvisational storytelling [106]. In contrast to the relatively simplistic formulaic improvisation strategies used in the majority of music improvisation systems, the systems presented in this work can learn domain-specific improvisational patterns directly from the users actions as well as use formalizations of general-purpose improvisational response strategies to act within regions of decision space where it has no prior experience. The emergency response decision support [100], Three Line Scene [26], and Party Quirks systems [56] were designed as cognitive models of the improvisational process. However, they maintain a static repository of encoded expert knowledge to use, and thus can only be used in significantly limited versions of the open-ended domains in which they were designed to improvise. They would also suffer from the improvisational action selection problem if their domains were expanded to be realistically open-ended. The proposed research aims to overcome this shortcoming in current approaches by having the system use creative arc negotiation to mitigate the improvisational action selection problem. Newer improvisational systems such as [104, 103, 106] offer exciting directions for addressing the improvisational action selection problem. However, all three systems fail to evaluate the creativity of their responses/offers before producing them, leading to reduced creative agency and more creative responsibility placed on their human collaborators for incorporating the system's outputs into the improvised performance.

2.2 Learning and Generating Actions For Movement Improv

The improvisational agents presented in this dissertation perform generative exploration of actions learned from demonstration in order for the agent to perform action space search while selecting actions improvisationally. The embodied knowledge that agents must learn in order to be able to respond successfully and expressively to human improvisational partners within movement improv varies based on the particular performance domain but could include 1) the set of gestures that can possibly be performed in the domain; 2) semantic knowledge about what the gestures 'mean', portray, or cause within the domain that is grounded in the agent's experience; 3) the set of causal, temporal, and aesthetic constraints, policies, patterns, or rules that in the agent's experience, allow it to sequence together the actions it knows about; and 4) other conceptual or procedural knowledge about its environment, performance, or other agents within the domain that it learns over time. The embodied agents presented (or referred to in prior work) in this research learn a subset of this knowledge to different degrees from their human collaborators with the iterative application of action learning from demonstration after every performance is completed (see section 3.3.4).

2.2.1 Learning from Demonstration (LfD) and Imitation Learning

Imitation learning and learning from demonstration (LfD) are both forms of observational learning. There are sometimes used interchangeably in the robotics and human-robot interaction (HRI) literature [107, 108]. The two terms can refer to different techniques for observational learning in the reinforcement learning (RL) community [53] however. In the latter context, LfD includes techniques like inverse reinforcement learning (IRL) where the reward function is learned from demonstration, and regular RL is used to learn a policy for optimizing that learned reward function [54], while imitation is restricted to techniques like behavioral cloning that learn both actions and policies from demonstration [55]. LfD

research in HRI includes learning compound action models from demonstration by learning the sequencing and structure of primitive actions that make up a compound action commonly using Gaussian mixture models, support vector machines, or hidden Markov models [107, 108]. Other research in this area focuses on the learning of task networks that encode a graph-based representation of the sequence of primitive actions and higher-level combinations of these primitive actions [109]. A recent approach also presents case-based imitation learning for transferring learned skills to new contexts [110]. LfD and imitation learning in both HRI and RL contexts are usually applied to heavily constrained and well-specified tasks due to their sample inefficiency and the necessity for providing enough demonstrations to cover the decision space. In contrast, the improvisational agents presented in this dissertation operate within open-ended, ill-defined problem domains where the lack of a well-specified reward function and the size of the action space make the preceding techniques difficult to apply.

2.2.2 Generation of Action Variants

The improvisational agents presented in this dissertation directly learn an explorable latent action space from a set of demonstrations. Research in gesture generation has explored related questions to this approach, where gestures are synthesized using different techniques. Older systems explored this problem used statistical sequencing of primitive gestural components, while more recent research has used direct synthesis using neural network approaches.

Gesture synthesis systems in disciplines such as choreography synthesis, robotics, and embodied conversational agents, try to create parameterized, natural, and expressive gestures by following a similar pipeline: input to gesture planner, selection by a statistical model, and modification by final component [111]. Generative choreography systems such as Ikeuchi [112] and Ofli, Erzin, Yemez, and Tekalp [113] used segmented music measures as a conditioning input to their generative choreography systems. The most statistically

likely candidate dance segments from a pre-authored database were then chosen based on the music inputs and combined to create smooth transitions. Embodied conversational agents create gestures from speech, text, or video clips. Mancini and Castellano [114] used video tracking and analysis to create an agent capable of mimicking detected expressivity. Kipp, Neff, Kipp, and Albrecht [115] focused on creating natural gestures in virtual agents by using g-units to create continuous flowing movements from gesture segments. Previous models of gestural creativity have been successful in mimicking tasks, but a deep generative model was chosen for action variant generation in the latter part of this work instead of a traditional statistical model because of the open-ended action space in the domain.

Gesture synthesis has made significant advances through deep generative models such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs). They have proven to be particularly useful for generating novel gestures and choreography with minimal feature engineering by hand. Augello, Cipolla, Infantino, Manfré, Pilato, and Vella [116] employed a vanilla VAE trained on a data set of human dance movements to generate robot dance movements. Similar work by Kiasari, Moirangthem, and Lee [117] focused on combining VAEs and GANs to produce sequences of stylized actions. Their model utilized latent variables from the autoencoder as input to the GAN's discriminator network, while the input to the GAN's generator network was conditioned using action labels and initial poses of the generated action sequences. The architecture presented in this work also seeks to control the mode of the generated data through conditioning but adds conditioning both at input and latent space sampling stages since we draw inference directly from the latent space (see section 3.4.3).

Recurrent Neural Networks (RNNs), notably Long Short-Term Memory (LSTM) networks, have also commonly been used for sequential motion generation. Researchers have exploited the hidden Markov model process underlying motion and choreography by using RNN models that combine distributed hidden states and non-linear dynamics. The results are evident in choreographic support [118, 119] and motion synthesis [120, 121]. The

approach presented in this work extends previous work by conditioning RNN-based generative models for gesture synthesis and preserving local/regional coherence by grouping multiple poses within temporal proximity.

2.3 Computational Formalizations of Tacit Knowledge

Improvisational practitioners rely on tacit knowledge of many kinds in order to successfully improvise performances in near real-time. These can include tacit knowledge in the forms of learned conceptual systems, frameworks, and vocabularies from within the domain for conceptualizing or reasoning about different aspects of the performance during the improvisation; learned constraints and rules specific to the performance company/troupe or within the improv theater game/activity being played/performed; and learned procedural knowledge about how to act in various situations, including procedural strategies for improvising in uncharted performance territory. This section describes formalizations of tacit knowledge that are implemented as procedurally encoded mappings and procedural strategies so that they can be applied across many different contexts rather than just being additional expert knowledge that needs to be repeatedly authored for every new improvisational agent.

2.3.1 Affordance Domain Knowledge

The research presented in this dissertation on improvisational agents for object-based gestural proto-narrative improvisation relies on the formalization of a learned procedurally encoded mapping between physical object attributes and the action variants within an agent's learned action space. This enables the agent's local action space exploration to be constrained to regions of the global action space that are 'afforded' by the physical attributes of that object. This section examines definitions of affordance and how they might apply to the current context.

The concept of *affordances* was first introduced by Gibson in his seminal work on ecological psychology [122, 123]. Over time, the term was adopted and adapted by designers

[124] and roboticists [125], among others (for a survey of perspectives on affordances, see [126]). The resulting blurring and adaptation of the meaning of the term 'affordance' have tailored the definition of the term to the needs of the individual field within which it is used.

The term was originally defined by Gibson in [122] as follows. "When the constant properties of constant object are perceived (the shape, size, color, texture, composition, motion, animation, and position relative to other objects), the observer can go on to detect their affordances." He elaborated on the meaning of the term in later writing [123] by stating that the "affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill. The verb to afford is found in the dictionary, but the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment." In this most commonly referenced description of Gibson concept of affordances, the relationship between animal and object/environment is clearer than his initial definition. Additionally, Gibson also specified the process by which affordances are perceived and utilized. An agent perceives an object's affordances by directly perceiving and recognizing its perceptual invariants to mean the presence or absence in that object of a particular affordance that enables it to perform the specific action corresponding to that affordance with that object, i.e., it enables the agent to use that object in that particular way. This means that the set of affordances are fixed for every object and are represented as binary presence-absence values for any particular action (of which there may be many thousands of actions).

Perspectives from other fields have also been useful for comparison with (and usage within) this work. Norman [124] provided a modified description of affordances as the relationship between the agent interacting with an object in its environment and the perceived actions that could be done with that object. In Norman [124]'s own words, "affordance refers to the perceived and actual properties of the thing, primarily those fundamental properties that determine just how the thing could possibly be used." This focuses on the

perception of the afforded actions in addition to the absolute affordances that are functionally available to the interacting agent regardless of the plausibility of discovery or usage. He later termed this 'perceived affordance' to distinguish it from Gibson's [123] usage of the term. Norman did eventually come to regret the casual eliding of both terms in field of design community and stated that, "[w]hen I get around to revising POET [[124]], I will make a global change, replacing all instances of the word 'affordance' with the phrase 'perceived affordance'."

Adapting later perspectives from ecological psychology from both Stoffregen [127], who stated that affordances are "properties of the animal-environment system" and that "they are emergent properties that do not inhere in either the environment or the animal", and Chemero [128], who stated that affordances are "relations between the abilities of organisms and features of the environment", Şahin, Çakmak, Doğar, Uğur, and Üçoluk [57] defined affordances as follows. "An affordance is an acquired relation between a certain effect and a (entity, behavior) tuple, such that when the agent applies the behavior on the entity, the effect is generated." Comparing this definition to previous definitions, the entity refers to the object and its properties, the behavior refers to the embodied capabilities of the agent that interacts with the entity, and the effect describes the outcome of some kind from performing an action on/with that entity. It is important to note that the relation between the effect and the entity-behavior tuple is not intrinsic, but acquired somehow, whether through previous interaction, explicit design, or some other process. This differentiates it significantly from Gibson's [123] original notion of affordances being part of the intrinsic nature of an object within the agent's environment. This definition, in combination with Norman's [124] earlier definition form the basis for the usage of the term 'affordance' in this work.

The concept of affordance in my work (defined in my research as "a learned tacit procedural mapping between the physical attributes of an object in the agent's environment and that agent's learned action space that partitions and controls access to that agent's action

space.") applies to an embodied agent situated within its environment alongside objects with which it has the embodied ability to interact. Affordance, in this context, is defined as a tacit learned mapping that is procedurally encoded between the physical attributes of an object in the agent's environment and that agent's learned action space that partitions and controls access to that agent's action space. This represents a relational mapping between the entity (i.e., the physical attributes of the object), its embodied capabilities, and the set of actions possible, similar to Şahin, Çakmak, Doğar, Uğur, and Üçoluk's [57] definition of affordances. However, since the agent's improvised actions are pantomimed, an action's effects are not physical but represented in terms of the agent's interpretation of what is being portrayed (or signified). The mapping, in turn, makes certain regions of the agent's action space (i.e., certain actions) more or less difficult to generate from (or even consider). In this way, it forms a hybrid interpolation between the absolute affordances (possible or impossible actions) that Gibson [123] describes and the perceived affordances (more or less easily perceived actions) that Norman [124] describes.

2.3.2 Procedural Improvisation Knowledge

Many tacit or learned genre conventions, structural rules, and procedural strategies exist in improv theater to facilitate successfully coherent and entertaining improvised performances from the performers despite the severity of the improvisational action selection problem for the ground-up embodied narrative improvisation seen in improvisational theater. For example, in long-form improv theater where improvised performances can rival the duration of a rehearsed play, there might be structural rules that improvisers have decided on beforehand such as fixed sequences for starting and ending scenes from multiple subplots as well as how to move between the parallel scenes from multiple subplots. This preserves the vastly open-ended and massively ill-defined nature of the long-form improv theater problem domain but provides some degree of constructive constraints for the improvisational process. Similarly, other improvisational art forms have their own conventions

and structures that facilitate coherent improvisation. The degree of structure can vary significantly, however, from trading solos for a fixed number of measures over a jazz standard to free jazz improvisation with far fewer constraints or the diversity in the degree of constraints between short-form and long-form improv theater. The following section presents examples of practice-based and computational formalizations of tacit procedural strategies from improv theater for constructing improvised narratives that are complementary to the improvisational response strategies and creative arc negotiation process presented in this dissertation.

Improvisational Conventions In Improv Theater

Conventions for improv theater include fundamental rules like, 'Yes, and...' which ensures that performers constructively build on top of creative offers from other players without stalling or halting the momentum of a scene, not being too clever when adding offers to the scene in order to support other performers' ability to react realistically to your offers, or respecting the bounds, physical reality, and constraints of an imaginary setting that has already been established by another performer [129]. Other conventions for certain kinds of narrative-driven improv theater act as high-level procedural strategies for prescriptively guiding action selection during the improvised performance (at least) at the high-level description of a practice-based framework. For example, Johnstone [130] describes the convention of establishing a *platform* (i.e. what is the setting for a scene, who are the characters in it, and what activity are they engaged in) as early as possible, then focusing on adding conflict to give the scene purpose, and then repeatedly acting together to create and resolve lesser conflicts or *tilting* the platform that has been established by adding new elements that reinterpret the scene and create a new narrative direction to explore for the group.

An alternative to Johnstone's high-level framework is the widely used practice-based framework from the Upright Citizen's Brigade (UCB) [131] involves finding the *game of the scene* after establishing the platform and then navigating the remainder of the scene

using three kinds of 'improvisational moves' — *raising the stakes*, *exploration*, and *top of intelligence* responses [132]. In this framework, raising the stakes involves performing actions that add conflict, drama, or some form of narrative incoherence that has to be resolved imminently. Exploration actions are those that justify some incoherence in the scene that arose from a previously performed raising the stakes action. Finally, top of intelligence actions are those that would form an established character's natural reactions to unusual or incoherent situations. According to this prescriptive framework, once the game of the scene has started, actions have to repeatedly raise the stakes, cause top of intelligence responses from other characters present, and then cause the unusual situation to be resolved or justified to some degree using exploration responses by one or more of the other characters until the scene ends with enough of a resolution of the raised stakes.

Improvisers also have to collaboratively construct a shared fictional reality in real-time, fluidly navigating ambiguity. Therefore, there are also tacit improvisational conventions around resolving ambiguity without halting the performance and the construction process. These conventions can be seen in various knowledge disparity games from short-form improv theater such as *Party Quirks*, where improvisers arrive as guests, one by one, at a host's improvised party with a quirk that is only known to themselves and the audience. The host is required to subtly investigate each guest's quirk and equally subtly guess what their quirk is without halting the improvised party or breaking character. The improvisers who act as guests gradually reveal increasingly obvious clues about their quirks until the host guesses correctly, but timed to provide the audience with a satisfyingly long time being the only ones (aside from the guest with the quirk) who know what the quirk is and understand the references or inside jokes pointing to that fact.

Improvisers use tacit procedural strategies and conventions even in straightforward situations like platform establishment, where ambiguity needs to be resolved as quickly as possible for the scene to proceed where an improviser A who is miming actions to convey that they are raking leaves outside their house might be understood by improviser B

to be sweeping the floor inside a cafe, there are clear conventions for navigating around each performer's beliefs about the ongoing improvisation. These strategies involve communicating what one improviser believes to be happening through actions related to the content of the scene, monitoring the other improviser's reactions to see if they are correct in their original beliefs, correcting misunderstandings using a set of repair strategies, and then covertly communicating that they (or the other improviser) have changed their beliefs to best support the scene moving forward. This tacit *shared mental model* negotiation process [133] is performed by experienced improvisers using procedural strategies so as to fluidly navigate the ambiguity of a shared fiction that is being simultaneously explored and constructed collaboratively while hiding the nature of the resulting cognitive divergences from the audience observing this negotiation process.

Computer Models of Improvisational Conventions In Improv Theater

Several procedural strategies and improvisational conventions have been modeled computationally in the literature by taking inspiration from the various practice-based frameworks as well as by studying how improvisers actually perform improv in laboratory experiments. O'Neill, Piplica, Fuller, and Magerko [26] describe a computational model for establishing the platform in a short form improv theater game called *Three Line Scene*. Brisson, Magerko, Brian, and Paiva, Ana [134] describe a computational model for finding the tilt and exploiting it for adding progression to a scene being improvised. Though not strictly from the domain of improv theater, Davis, Comerford, Hsiao, Jacob, and Magerko [3] present the (partly computationally-implemented, partly conceptual) enactive model of playing pretend (which can perhaps be understood as a less structured or stylized form of everyday non-expert improvised narrative in comparison to improv theater). Davis, Comerford, Hsiao, Jacob, and Magerko's particular model presents a marked similarity to the UCB framework for finding the game of the scene. Magerko, Dohogne, and DeLeon [101] described the improvisational strategies for navigating knowledge disparity games in the

Party Quirks short form improv theater game and Hodhod, Piplica, and Magerko [135] described an architecture for formally modelling the shared mental model negotiation process in improv theater games. Additionally, Martin, Harrison, and Riedl [106] proposed a system for open-world improvisation using plot graphs that included a set of strategies for handling user actions as they were classified under constituent, consistent, and exceptional branches as a way to perform disembodied improv theater in the context of text-based narrative improvisation.

The preceding computational models of procedural strategies for navigating various components of the narrative improvisation task in improv theater heavily rely on rich, pre-authored domain knowledge in order to function effectively. They are severely impacted by the limitations of content authoring and need to address that problem before they can be used to address the improvisational action selection problem. Martin, Harrison, and Riedl [106] use plot graph learning from crowdsourced knowledge to avoid this problem in their text-based improv domain. However, the problem remains challenging to address when working with embodied content knowledge to enact the crowdsourced natural language knowledge. As the semantic complexity and formal structure of the knowledge learned by improvisational agents such as the ones presented in this dissertation increase to the point where they can be used by these models, the approaches presented above would greatly help to add structure and constraints to the ill-defined nature of the improv theater domain. Therefore, they remain complementary to the domain-independent improvisational response strategies and the creative arc negotiation process presented in this dissertation.

2.4 Computational Models for Evaluating Creativity

There have historically been a large number of theories, models, and definitions for understanding (and subsequently evaluating) creativity from diverse research fields ranging from media and cultural studies to psychology and cognitive science to artificial intelligence and computational creativity. One way to organize this set of parallel ideas usefully is to use the

4P taxonomy from Rhodes [136] that was later introduced to the field of computational creativity by Jordanous [137]. The 4P taxonomy organizes theories, models, and definitions of creativity into the four categories of *person* or *producer*, *product*, *process*, and *press*. Models of creativity that deal with person or producer focus on intrinsic characteristics that make that person or producer creative (e.g., psychometrics for creativity [61, 138] or case studies of creative people [139]). Product models focus on the creative artifact produced as a part or result of creativity, while process models focus on modeling the creative process itself. Press models focus on how the creative person/producer, process, or product affects the culture, environment, or society within which it exists.

Several perspectives on creativity evaluation also deal with multiple aspects of the 4P taxonomy simultaneously (e.g. Colton, Jordanous, Colton, Charnley, and Pease). Colton's creativity tripod [9] argues for evaluating a computationally creative system on the basis of its skill, imagination, and appreciation of the creative medium seems on the surface to be a producer/person model of creativity evaluation by the traits or characteristics of the system. However, this work is intended in a way that addresses both person/producer and press models of creativity evaluation.Jordanous' Standardized Procedure for Evaluating Creative Systems (SPECS) methodology [11] describes 14 criteria obtained through the analysis of creativity research corpora that are suggested for use by researchers in creating their own working definitions of creativity in order to rigorously evaluate the creativity of their system according to that specific working definition. Colton, Charnley, and Pease's FACE model [10] evaluates computational creativity systems for creative concept invention, expression of the concept as an artifact, aesthetic evaluation of the artifact, and the framing of the artifact to the public.

The research presented in this dissertation uses a product definition of creativity (see section 1.1), especially as part of the agent's creativity evaluation models that are operationalized in this work (see section 3.3.6). However, the discussions of improvisation as a process (see section 2.3.2), as well as the process of creative arc negotiation described in

this work (see section 3.3.3), focus on process-based aspects of creativity. Therefore, this section focuses on the process and product perspectives on creativity and will ignore the person/producer and press perspectives. A part of the evaluation sections for the interactive installation (see section 3.5) presented in the dissertation references press-based measures of creativity, so the press perspective will be referenced as needed in context. For more detail on all four categories, a detailed survey of models for evaluating creativity and how they impact the methodology of computational creativity research can be seen in Lamb, Brown, and Clarke [140].

2.4.1 Product Models of Creativity

Product models of creativity focus on the different potential qualities of an artifact that enable it to be called creative to some degree. The most famous of these models is Boden's [8] model of creativity for artifacts that considers an artifact creative if it possesses various kinds of novelty and value while evoking surprise in an experiencing entity. This conceptual model was operationalized by Maher [141] to measure the novelty, surprise, and value for artifacts resulting from design creativity. Ritchie [142] provide a parallel evaluation framework for creative artifacts that features the additional criteria of typicality (or conformity) to the expectations for artifacts in a domain. This added criterion emphasizes the desirability of both novelty as well as typicality depending on the context. Since it provides mathematical models and computational functions for evaluating creativity, the work presented in this dissertation extends Maher [141] to provide computational models for evaluating the novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value) of perceived and generated actions within improvisational domains. Additionally, the systems presented in this dissertation incorporate a central idea that aiming to maximize novelty and unexpectedness are not always the most important aspects of a temporally-extended session of improvisational creativity. This is in keeping with both Ritchie's typicality and Perišić, Štorga, and Gero's situated novelty.

Pease, Winterstein, and Colton discuss how modified Turing Test evaluation methodology emphasis computational pastiche and 'window dressing' rather than actual creativity in generated creative artifacts. They recommend framing information as additional artifacts that are needed for a computationally creative system as such. Lamb, Brown, and Clarke [140] also agree and recommend it as methodology only if verisimilitude to human artifacts is a true requirement for creativity in a system's generated artifacts. Taking this advice to heart, the modified Turing Test is only used to assess the properties of outputs from the improvisational agents when verisimilitude is the point of the evaluation.

2.4.2 Process Models of Creativity

Process models of creativity describe different aspects of the creative process itself, rather than focusing on the outputs of that process necessarily. Newell, Shaw, and Simon [7] present a search-based process for creative problem solving that involves searching for solutions to a creative problem that are novel and have value (similar to product theories), but also focus on the rejection of previous assumptions, demonstrating persistence towards a goal, and the development of the problem specification itself over the course of the search process. Boden [8] categorizes the different processes of generating creative artifacts (process in relationship to product creativity) into three categories — *combinatorial*, *exploratory*, and *transformational* creativity. All three forms of creativity are described in relation to the conceptual space of the creative domain. Combinatorial creativity involves combining elements of a single (or multiple) conceptual space(s) together in order to generate creative artifacts. Exploratory creativity is a search through a conceptual space to discover different creative artifacts within the bounds of that space. Transformational creativity is the transformation of the 'rules' or bounds of a conceptual space to be able to generate creative artifacts that couldn't be generated within the bounds of that space previously, even with the most exhaustive exploration. Combinatorial creativity includes processes such as *conceptual blending* [145] (where elements of two input concept spaces are selectively

mapped together and combined into an output conceptual blend space), *analogy* [146] and *metaphor* [147] (different processes for selectively mapping and transferring elements from a source space into a target space to create new conceptual outputs), *conceptual expansion* [148] (combining concepts from different spaces using mathematical filters to select and modify how their elements are transferred into the expanded space), and other forms of conceptual combination in the literature. Exploratory and transformational creativity were further operationalized beyond Boden's original abstracted description by Wiggins [149] to be defined as formal search within a conceptual space (exploratory creativity) as well as meta-search of conceptual spaces themselves (transformational creativity). The research presented in this dissertation intentionally includes techniques for computational creativity that covers combinatorial and exploratory creativity. However, any transformational creativity exhibited by the system is not intentional.

Process theories of creativity have also included prescriptive or descriptive models of different stages that are involved in the creative process. Wallas' idealized four-stage model of creativity includes preparation (information gathering), incubation (considering the problem and perhaps abandoning a conscious search for a solution), insight (spontaneous awareness of a solution to the problem that was potentially abandoned), and verification (evaluation to see whether the idea works and then modification or development as needed). Others have included additional stages such as intimation [151], evaluation [152], and interactions between explicit and implicit reasoning/knowledge during these stages [153]. Other stage models include Finke, Ward, and Smith's Geneplore model [154], Johnson-Laird's NONCE model of improvisational creativity [155], and Perez and Sharples' exploration-reflection (ER) model (originally applied to creative writing). The Geneplore model consists of the exploration and evaluation of pre-inventive structures generated through synthesis, transformation, and exemplar retrieval as a description of the creative process. The NONCE model for jazz improvisation involves performing both knowledge or constraint-guided generation (neo-Lamarckian approach) and explicit constraint-

based evaluation after generation (neo-Darwinian approach). The ER model involves alternating sequences of exploration (or generation) of content fragments, followed by a reflective evaluation of the generated fragments until a satisfactory creative solution is created.

The process models described in this section all commonly perform some repeated or cyclical stages of generation, followed by evaluation. The research presented in this work presents creative decision making on the part of an improvisational agent architecture that can be described as a process model for improvisational creativity. The creative arc negotiation performed in this research follows a cyclical generate and evaluate process, where the agent's generation process is constrained by the temporal bounds of the performance, the agent's given creative arc, and its set of improvisational response strategies. The agent directly evaluates each of the candidate responses it generates for novelty, unexpectedness, and quality fit to a given creative arc target point. In this way, it is close to the hybrid between neo-Lamarckian and neo-Darwinian approaches that is advocated for and theoretically described in [155].

2.5 Creativity and Intrinsically Motivated Agents

Models of creativity can also serve as motivational drives for agents. For example, curiosity, defined here as an intrinsic motivation to discover novel percepts, experiences, explanations, or knowledge [156, 157], is one of the intrinsic motivational drives that can be used to control learning algorithms. This is exemplified by curiosity-driven reinforcement learning (RL) (e.g., [158]), where curiosity is used to modulate an agent's learning process. Schmidhuber also describes curiosity as the intrinsic reward mechanism that enables their agent to learn in the absence of external reward functions in domains such as art and music. Other intrinsic motivation functions can also be used to modulate agent behavior. For example, Guckelsberger, Salge, and Colton presents intrinsically motivated agents for co-creative contexts based on coupled empowerment maximization. This drive is a multi-agent generalization of empowerment maximization [161], which is the potential for an

agent to influence or control the outcome in the future given the current situation. Creative arc selection and negotiation are presented in the latter part of this dissertation and can be conceptualized as intrinsically-motivated search-based solution generation. In this case, the target trajectory is the specified creative arc, and the intrinsic motivation is the drive to follow that arc as best as possible within agent turn time limits. There is a far greater challenge in the current context compared to the previous approaches due to the near real-time nature of the improvisational domain.

Evolutionary computing is another area where product-based aspects of creativity serve as intrinsic and unique motivation for solution search. Objective search [162] is the default configuration for evolutionary computing for ordinary or search-based creativity applications and corresponds to a search for high-quality solutions. Objective search involves performing exploration of the solution space using a population of evolving candidates. However, in the recent past, novelty search [163] and surprise search [164] have seen much success in finding solutions more rapidly or robustly than objective search. This is particularly the case where finding globally optimal solutions requires the algorithm to traverse deceptive paths through the search space [164]. Most of these systems have yet to investigate strategies for choosing between novelty and surprise or combining them as the need may be. It is also technically possible to simulate a similar novelty seeking (curious) agent in the proposed decision-making model by providing it with a creative arc that has a maximal novelty dimension throughout its trajectory, along with setting the agent to ignore (or alternatively, to accept any value in) the surprise and value dimensions of the creative space. This also enables the agent to perform surprise search [164] by providing the system with a creative arc that maximizes the surprise dimension while setting the agent to ignore (or alternatively, to accept any value in) the other two dimensions. The creative arc negotiation system can also simulate other hybrid search agents [165].

The creative arc following agent presented in the latter part of this work differs from the various intrinsically motivated agents mentioned above in the following two ways. Firstly,

the former directly optimize novelty, surprise, and value dimensions (among others) while the latter tries to optimize a given meta-level function composed of those same dimensions in order to follow a creative arc. This is in contrast to the other techniques which encourage always generating the most creative response. For example, a creative arc that starts with low novelty and progresses to some peak novelty value before descending again might be more valuable to an improvisational partner than an agent that tries to do the most novel action it possibly can every single turn. Secondly, in the former cases there is often a final output to the search process (when search is stopped eventually) that is evaluated to assess the effectiveness and quality of the optimization technique, while in the latter case, the agent's creative artifacts are experienced by the agent's improvisational partner (and a potential audience) all throughout the creative arc making the improvised journey itself the main creative artifact that is output for and assessed by the audience, not necessarily any individual action generated by the agent along the way. The agent evaluates the creativity of perceived and generated each action over the entire course of the performance, though.

2.6 Interactive Narrative and Drama/Experience Management

There is a natural fit between the eventual goal for this research as a path towards embodied narrative improvisation and work in drama or experience management [166, 167] within interactive narrative [168] research. Both seek to enable co-creation of entertaining (or desirable) user experiences for participants (and potentially for an audience as well). The research presented in the latter half of this dissertation on creative arc negotiation was at least partly inspired by interactive narrative systems such as Mateas and Stern's *Façade* [169], Porteous, Teutenberg, Pizzi, and Cavazza's Merchant of Venice [170], and Magerko's *Haunt II* [171] as well as Riedl, Stern, Dini, and Alderman's *Automated Story Director* (ASD) [172]. Façade [169] uses annotated story fragments called *beats* that are sequenced in response to natural language user inputs with a reactive planner to generate interactive narrative experiences for the player that possess a clear dramatic arc [173].

Similarly to Façade, the creative arc negotiation process presented in this work attempts to structure the improvised performance according to its creative arc. A creative arc can emulate arcs over a dramatic tension space by adding that as a component to the agent's quality heuristics and setting the agent to ignore novelty and unexpectedness in candidate actions. However, the creative arc negotiation can also use additional criteria such as novelty and unexpectedness, to create player experiences that may only be accidentally possible for an interactive narrative system such as Façade. Porteous, Teutenberg, Pizzi, and Cavazza's work describes a visual programming method for drawing dramatic arcs in order to guide a planning-based interactive narrative experience of the Merchant of Venice. This is a potentially useful idea that my research could incorporate in the future to enable the personalization of creative arcs to players or ease the authoring process for non-expert experience designers. Experience or drama management systems in interactive narrative like Haunt II [171] and ASD [172] predominantly manage the tension between authorial intent and player agency in the interactive to preserve the coherence of player or user experiences (see [174] for a survey of the interactive narrative space and [167] for a survey of drama management techniques). The techniques presented for managing the coherence of the user's experience in [167] could be incorporated in the future, once the improvised performances are closer to narrative than proto-narrative. However, the most significant problem with using these techniques at the moment is the extreme level of content pre-authoring required to create structured knowledge for use in these experience management techniques, which needs to be addressed before they can be integrated into open-ended movement improv systems.

CHAPTER 3
THE ROBOT IMPROV CIRCUS

My thesis statement for this dissertation states that "embodied agents that address the improvisational action selection problem using 'creative arc negotiation' increase perceptions of enjoyment, agent creativity, and coherence in both observers and participants while performing movement improv with non-experts." In order to investigate the specific claims within that thesis statement, this chapter uses the following outline. I first introduce the Props game problem domain and the Robot Improv Circus installation within which this problem was studied. I then describe the general solution approach for addressing the improvisational action selection problem using creative arc negotiation and how that is implemented within an embodied improvisational agent architecture called CARNIVAL for enabling embodied agents to perform movement improv with non-experts in the Props game domain. I then discuss several experiments that aim to validate architectural components and systematically evaluate the extent to which creative arc negotiation addresses the improvisational action selection problem and improves participant and audience perceptions of enjoyment, agent creativity, and coherence as stated in my thesis statement.

3.1 The Props Game Domain

My research investigating the claims in my thesis statement and understanding how to build embodied improvisational agents to perform movement improv with non-experts and address the improvisational action selection problem was situated within *object-based gestural proto-narrative improvisation*. This domain is defined as proto-narrative improvisation performed using gestural interaction with objects in the agent's environment (see section 1.3). This form of improvisation is exemplified by the popular short-form improv theater game — the 'Props' game. The Props game involves improvised movement-based interac-

Figure 3.1: Two actors playing the Props game from the popular TV show, "Whose Line Is It Anyway?" [175]

tions between two or more participants using ambiguous props given to them at the start of a game round to perform recognizable comedic vignettes or singular actions pantomiming the use of the abstract prop as a real-world or fictional object. For example, when presented with a prop shaped like a long, thin pole with a small sphere on one end, the first performer pretending to use it like a baton and twirling it about like a bandleader for a marching band, then the other performer pretending to play a drum solo using it as a long drumstick, and continuing on with different props over different rounds of the game. Many different variants of this game exist in the improvisational theater community with different degrees (and kinds) of connectedness between performer turns, but the previous definition is the version used in this research.

The Props game domain was chosen because the challenging nature of the *improvisational action selection problem* (described in section 3.3) had been highlighted in prior work [2] and the Props game domain allowed me to focus on that specific problem. Addi-

tionally, in comparison to domains closer to unconstrained embodied narrative improvisation, the focus on one-shot actions or short vignettes through prop-based improvisational interactions represented a simplification of higher-level semantic reasoning and longer-time scale temporal reasoning for creating a virtual improviser in the domain. Beyond improv theater, solutions to the Props game problem could potentially be adapted to allow embodied agents to gain new knowledge about unfamiliar objects through interaction. For example, with the addition of additional evaluation heuristics and goal-oriented reasoning, the technical approach used in this work could enable agents to learn to use unfamiliar objects in unfamiliar scenarios according to familiar human norms/customs or use unfamiliar objects for a specific task (such as improvising a digging tool for disaster recovery). In summary, the motivation for selecting the Props game was that it allowed me to focus on the improvisational action selection problem while simplifying the complexity of the problem to be more feasibly addressable (in comparison to unconstrained embodied narrative improvisation) and enabling the future extension of this work to other important applications.

The agent's actions within the Props game domain consist of gestural content and semantic content. The gestural content represents the positions and orientations of key skeletal points of a body over time. The semantic content consists of the semantic interpretation of the gestural content of an action in terms of the English verb describing the pretend action being pantomimed as well as the English noun describing the pretend object being signified with that pretend action. Since the semantic content represents an interpretation of the gestural content, it constrains the space of valid gestural content to those that can be mapped to interpretable semantic concepts of pretend objects and pretend actions. Thus the Props game represents a tangible increase in the complexity of the improvisational domain over prior work in purely gestural proto-narrative improvisation [50, 2, 16] due to the increased semantic constraints imposed over the agent's generated gestures in order for them to be interpreted successfully. However, it also does make it easier for the agent

to communicate with the human user about how to interpret improvisationally generated actions.

The added semantic constraints in the domain prevent certain computational techniques (e.g., many forms of data augmentation) from being applied since naive modifications to the gestural content of an action might not align with the interpretive semantics of the augmented gestural content, and it is not possible to know about such a change in advance. Additionally, the semantics necessitate reasoning over additional modalities and content for communication with the human collaborator. This requires an expansion in the scope of the improvisation being performed. In contrast to the challenges previously described, the agent also obtains the opportunity for added clarity through communicating the semantics of the action. For example, the agent can indicate what both the intended pretend action and pretend object of a generated action were.

The current domain was studied within an interactive virtual reality (VR) installation where the Props game could be played between an embodied virtual agent and a human scene partner. This VR installation is called The Robot Improv Circus [176] and is being developed using an iterative design process. The installation has human users play the Props game as a humanoid robot with their humanoid robot stage partner on the main stage of an all robot circus. The experience was designed to leverage existing user expectations about setting and experience for circuses in contrast to improv theater performance venues. For instance, common expectations in the USA for circuses that can be leveraged to quickly situate a circus experience for many people include that they are held in colorful big top circus tents with specific circus music and other thematic elements. In contrast, improv theater possesses less specific and less commonly held cultural expectations that can be leveraged by the installation directly. Additionally, the denizens of the virtual world (human user included) were designed to look like humanoid robots rather than human characters in order to manage expectations about the realism and verisimilitude of the computational improviser's actions and behavior.

Figure 3.2: The user miming an action with a prop

3.2 The Robot Improv Circus Installation

The Robot Improv Circus [176] is a primarily single participant VR installation for people to play the Props game with a virtual agent. While the participant is performing in VR, the audience views the improvised performance from outside the installation through large screens that form portals into the virtual world. The experience takes place on the virtual stage of a robot circus, where improv is the main event. Participants take turns with the virtual agent to mime pretend actions using abstract props as a real-world or fictional object in imaginative ways in order to create an object-based gestural proto-narrative with the agent. For example, when presented with a prop that looks like a long, thin cylinder with a flat disk on one end, one player might pretend to use it like a katana and pantomime slashing at the air repeatedly. The other agent might then use that prop as a 'bo' (long staff) to pantomime blocking sword slashes or other actions.

The VR experience consists of a trial round followed by a small number of game rounds. Each performer is given a new prop every round, and each round consists of five to seven turns. The goal of each round is to create a proto-narrative by taking turns miming actions

Figure 3.3: A view of the virtual agent miming an action using a prop in the Robot Improv Circus VR installation

with the prop. Performers hit a virtual buzzer after enacting their actions to signal the end of their turn.

As an example, after receiving a prop shaped like a flattened cuboid, the VR user might pretend that the prop is a stovepipe hat and mime putting it on. She then hits the buzzer to end her turn. The same prop then appears in front of the agent who pretends to comb its hair using it as a comb. The agent speaks and displays a speech bubble that reads, "I am combing with a comb" (like in fig. 3.3). The speech and speech bubbles were added to encourage dialogue and increase the recognizability of the generated mimed actions after initial validation experiments showed a clear need for improving that aspect of the generated actions.

The Robot Improv Circus is exhibited in a circus tent (see fig. 3.4). The form and decor of the installation were designed to evoke a familiar circus aesthetic with circus flags, themed posters promoting the robot improv circus, and the VR experience itself housed within the circus tent. The large, colorfully decorated circus tent also seeks to create a commanding visual presence for the installation [177] to draw people to the installation.

Figure 3.4: A VR user experiences the Robot Improv Circus in the installation tent

Figure 3.5: A screen displays a view from the virtual audience to human audience members watching from outside the installation.

The installation has two large displays outside the circus tent that act as portals for a real-world human audience to glimpse the virtual circus stage and the performance being improvised inside VR. They can watch, applaud, and provide positive feedback to participants in VR through their physical activity. The activity of the human audience is captured using a video camera and pose extraction from the video feed [178], which then triggers different kinds of supportive visual feedback in the virtual world according to the amount of movement in the video frame over time. Their feedback appears in the virtual world above the virtual audience's heads as floating emoji (thumbs-up symbols, smiley faces, clown faces, and hearts) rising up from the robot audience.

3.3 The Improvisational Action Selection Problem

The primary focus of my research in this dissertation is to investigate how addressing the *improvisational action selection problem* for embodied improvisational agents that perform movement improv with non-experts in the Robot Improv Circus affects both participant as well as audience perceptions of enjoyment, agent creativity, and coherence. In order to do so, the embodied improvisational agents instantiated in this research must be successfully able to address the improvisational action selection problem. The improvisational action selection problem (see Section 1.2.1) refers to the challenging nature of near real-time action selection within improvisational domains that have open-ended action spaces as well as ill-defined goal spaces. Addressing the improvisational action selection problem requires a balance in the reasoning process that avoids decision paralysis, incoherent behavior, responses that are less diverse (or too similar), and a lack of qualitative impact on the user's experience.

3.3.1 Technical Need

Previous approaches to addressing the improvisational action selection problem have focused on various techniques to one or a few aspects of the problem. These include con-

straining the temporal responsiveness of the system [179], constraining the action space [26], simplifying or enumerating a constrained formalization of the goal space of the domain [100], or using simplified stochastic action selection [97]. My research attempts to address the different components of the improvisational action selection problem together and thus does not use any of the former simplifications.

The improvisational action selection problem has multiple interacting factors that make it particularly challenging to address. Firstly, the open-endedness of the action space makes near real-time performance difficult. Secondly, constraining the action space to make performance more responsive decreases the expressivity, flexibility, and diversity of possible agent responses [26]. Thirdly, if a simpler or more stochastic action selection mechanism is used to improve responsiveness, the agent's behavior seems incoherent over time [2]. Finally, the lack of a well-defined set of goals for the domain prevents the agent from confidently preferring one action over the other, leading to less diverse user experiences across the different versions of the system [180].

The ill-defined goal space for the improvisational problem domain makes it especially difficult for commonly used techniques like reinforcement learning (RL) [53], inverse RL (IRL) [54], or behavioral cloning (BC) [55] to be used easily. RL involves learning a policy for selecting actions to maximize reward over time and would be difficult to apply due to the lack of a well-defined reward function. IRL involves the learning of a reward function from observation and then using RL to solve the learned reward function. BC is the process of learning and generalizing action sequences used by experts in demonstrations to a new task. IRL and BC are difficult to apply due to the open-endedness of the action space alongside the ill-defined goal space. More specifically, this is due to the sample inefficiency of IRL and the relatively poor performance of BC in unobserved regions of the problem space.

3.3.2 Solution Approach

The approach used in my research attempts to address as many of the interacting components of the improvisational action selection problem simultaneously as possible. It is desirable for the agent to be able to improvise in near real-time within the open-ended action space to produce expressive responses that form coherent behavior over time while showing perceivably diverse behavior with changes to its action selection within the ill-defined domain. The following is the technical approach used to achieve these results in the improvisational agent presented in this research.

The computational approach to embodied improvisation used in this work aims to produce responsive improvisational behavior that is coherent over time with a demonstrably perceivable (or identifiable) impact on user experience (in terms of perceptions of enjoyment, agent creativity, and coherence) across different versions of the system. In order to generate improvisational behavior that satisfies these properties, this research presents a process called ***creative arc negotiation*****, where the agent performs stochastic, interruptible, strategy-guided action space search to follow a given *creative arc* through its *creative space* of novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value)** (see section 1.4). The agent implements creative arc negotiation using the following components. **Affordance-based action variant generation enables the agent to perform conditional parameterized generation of action variants based on the physical attributes of objects** that are given to it during improvisation as a way to search its learned action space. Adapted from prior work [50] and formalized from human improvisational practice, the agent uses **improvisational reasoning strategies to guide action space search** while negotiating a creative arc. The agent also **computationally evaluates the novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value) of perceived actions and generated candidate responses** to localize them within the agent's creative space. The following subsections describe the creative arc negotiation process and the conceptual details of the three components in more detail, however,

the exact implementation details for each component are described in the corresponding subsections of the CARNIVAL architecture (see section 3.4).

3.3.3 Creative Arc Negotiation (RQ4)

The agent presented in this research performs creative arc negotiation. *Creative arc negotiation* is the process of selecting actions over time to best follow a given target trajectory or 'creative arc' through an agent's 'creative space.' The agent aggregates both the human's last action and the agent's candidate actions before comparing it to the target creative arc in order to select the agent's next action. Creative arc negotiation aims to provide the user with a temporally evolving experience that appears coherent over time and qualitatively different across different creative arcs (or without a creative arc guiding action selection).

The working definition of creativity used in this system is an extension of Boden's definition of creativity focusing on the *novelty*, *surprise*, and *value* of perceived or generated artifacts (see section 1.1). A multidimensional model of creativity is used here with the artifact localized to a point in the space of novelty, surprise, and value. In order to avoid some of the deeply overloaded semantics of the terms 'surprise' and 'value,' in practice, the agent computationally evaluates candidate responses using heuristics of *novelty*, *unexpectedness* (as a measure of surprise) and *quality* (as a measure of value) instead. *Creative arcs* are, therefore, continuous trajectories through this three-dimensional *creative space* that an agent follows over the course of a temporally-extended improvised performance. An example of a creative arc is illustrated in Figure 3.6

Creative arc negotiation selects actions using a given trajectory over time (the creative arc) within the creative space rather than always attempting to choose a maximally creative action. There are several reasons for this. Firstly, the process was inspired by (and generalized from) practice-based conventions across several different creative domains about the use of arcs and trajectories to structure experiences. For example, tension in musical composition [181] and improvisation [182] (even pitch for simple *cantus firmi* [183]) follows

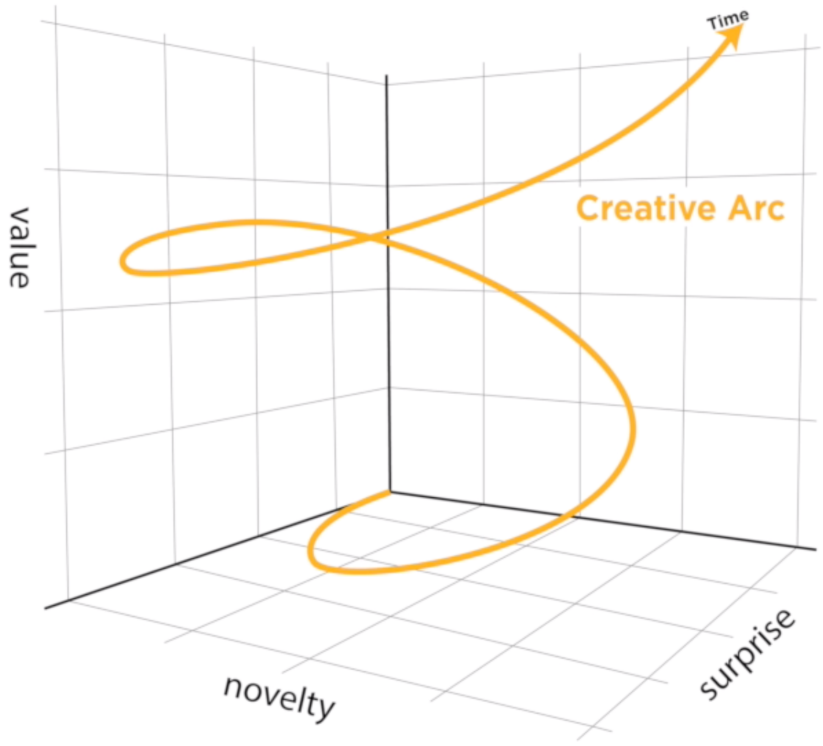

Figure 3.6: An example creative arc.

well-defined arcs. Similarly, dramatic arcs in several forms of narrative at a high-level rely on variations of a familiar trajectory involving rising to a climactic point and then falling to a resolution to deliver their affective payload. In visual art, as well, artists are often taught to compose their subjects so as to encourage a viewer's eye to move across the entire composition in smooth arcs and trajectories drawn by the visual forms and expectations of the composition. Secondly, the novelty and surprise components of creative space demonstrate 'inverted U' characteristics against perceptual arousal (or preference) [184]. Thus having a constant value at a maximum could be negative overall.

The improvisational agent is required to perform action selection in an open-ended action space in near real-time. Therefore, creative arc negotiation is strategy-guided as an optimization using strategies formalized from human improvisers that anchor the search in

the current improvisational context while searching desirable regions of the search space. Creative arc negotiation is implemented as an interruptible stochastic search where the agent returns the 'best' solution until a) the point of interruption in its search or b) it successfully finds an action within ϵ of the target creative space point as a further optimization.

3.3.4 Affordance-based Action Variant Generation (RQ1)

The improvisational agent learns the action space it can use for improvisation within the Props game from training data representing non-expert demonstrations of pantomimed actions using props as real-world or fictional objects. A conditional variational autoencoder [185] architecture was used to learn the distribution of the data set in its latent space in order for the agent to be able to search or explore that action space (including unseen action variants interpolated between the demonstrated examples). Autoencoders [186], in general, learn a non-linear dimensionality-reduced representation of a given training data set in their latent space. However, variational autoencoders [187] warp the latent space to allow smooth interpolations between trained data in the model's latent space. Conditional variational autoencoders (CVAE) are variational autoencoders that are conditioned on additional features that can be used to partition the latent space and condition generation from the model's latent space. The CVAE used in this research was conditioned on the physical attributes of props (see Section 3.4.3) used to enact the pantomimed actions from the data set. This allowed the agent to both restrict action variant generation to appropriate props but also to generalize learned actions to props with similar physical attributes (see section 3.4.3 for more detail).

I refer to the process of generating action variants from the agent's learned action space as *affordance-based action generation*. I define affordance in this work as "a learned tacit procedural mapping between the physical attributes of an object in the agent's environment and that agent's learned action space that partitions and controls access to that agent's action space." As described in section 2.3.1, this definition represents a relational mapping

between the entity (i.e., the physical attributes of the object), its embodied capabilities, and the set of actions possible with that entity, similar to Şahin, Çakmak, Doğar, Uğur, and Üçoluk's [57] definition of affordances. This learned mapping is encoded in the CVAE model as embodied tacit knowledge; therefore, the action generation proceeds using the learned model of affordance as defined above. The learned structure of (and relative distribution of action classes in) the model's latent space make(s) certain regions of the agent's action space (i.e., certain action classes) more or less difficult to generate from. It forms a hybrid interpolation between the absolute affordances (possible vs. impossible actions) that Gibson [123] describes and the perceived affordances (more vs. less easily perceived actions) that Norman [124] describes.

The agent learns how to generate action variants for generating candidate actions and searching its action space by training on a data set of human actions pantomiming the use of props as real-world or fictional objects. The data, once collected and annotated (see section 3.3.4), contains the physical attributes of the actual prop, the mimed pretend action, and the intended pretend object. The gestural content of the percept along with this interpreted information jointly form a perceived action in the context of the improvised Props game performance. The representation of the physical attributes of the prop is discussed in section 3.3.4. The combination of the pretend action and pretend object represents the semantics of this action (to a degree) based on the distributional hypothesis [188] that "a word is characterized by the company it keeps" (especially with their vector representation presented in 3.4.1).

Object Physical Attribute Representation

A key requirement for playing the Props game is that the given abstract (or unfamiliar) object/prop has to be interpreted into a real-world or fictional object and then used in such a way that signifies what object it is being pretended (or imagined) to be through pantomimed pretend actions (e.g., pretending a big sphere is a beach ball and pantomiming the

act of playing volleyball with it). In order to do this pretense computationally, the agent needs to be able to map the abstract or unfamiliar props to real-world or fictional pretend objects and vice versa. Annotated data from human demonstrations provide the agent with knowledge about what pretend objects and pretend actions are possible in the world and what embodied knowledge in the form of gestures actually corresponds to using a prop as a specific pretend object through a specific pantomimed pretend action. In order to extend and augment the agent's ability to perform affordance-based action variant generation an abstracted representation of object physical attributes was formalized. The process for developing this representation and the actual representation itself are presented below.

The agent's representation of objects in terms of their physical attributes was arrived at by reviewing affordance representation schema from robotics research [189, 190, 191, 192]. Varadarajan and Vincze [189] was chosen due to the two-stage process it introduced for mapping object features to primitive actions for agents to use [193]. Since only a severely limited version of their AffNet 2.0 database was available, the limited set of feature descriptors used in their work were filtered for feasibility, extended for coverage, and adapted for suitability with the Props game domain. Their process, however, was not feasible to use in this system due to its dependence on a fixed set of action primitives for learning a mapping.

The representation that was developed is as follows. A given prop is represented as a fixed-length feature vector. The feature values are obtained by decomposing the prop into a set of parts or components. This is done by comparing them to a fixed set of shape primitives and a fixed set of operations or deformations that could be applied to it. Since the focus of this research is not on the automatic segmentation of the prop into parts, this admittedly subjective decision was considered sufficient for hand annotation. The chosen individual parts of the object are then coded/parsed to obtain a set of binary physical attributes features representing whether or not that feature is present/applicable to that part.

The set of physical attributes features developed in this project includes a part's shape

primitive, size, thickness, flatness, concavity, taper, rigidity, curvature, hole size, and whether a digit/symbol is signified. As mentioned earlier, the feature set was chosen by extending the geometric mapping features from affordance representation ontologies such as [189]. After annotating the physical attributes feature values for each part, the features for each part are then aggregated by summing them together and normalizing them using the maximum count for any feature in the data set. This encoded value represents the normalized counts of each physical attribute feature for the prop across all parts and is a fixed-length vector representation of the object. For example, a barbell-shaped prop might be two flattened spheres connected by a long, thin cylinder. The process for encoding the physical attributes of objects could be automated in future extensions of this work using computer vision tools.

The extended set of features developed in this work for object representation was used to annotate a set of objects used by performers while playing the Props game on the *Whose Line Is It Anyway?* [175] TV show. A subset of eighty props was then hand-annotated using the developed feature schema out of the total number of props used on the show. Out of the eighty props that were annotated, twenty props were then selected for use in the Robot Improv Circus according to the following process. Fifteen props were chosen that had the highest aggregated individual feature counts as a rough measure of the total number of actions that could be performed with them. The remaining five props were chosen to accommodate props that had features which were absent in from the already chosen fifteen in order to boost the diversity of actions and actions possible with the set of props. The decision to choose props with high feature counts was made to potentially enable users to perform a larger and more diverse set of actions with these props. This decision itself was not specifically evaluated, but each prop (as well as the total set of chosen props) has empirically and qualitatively been shown to generate many different pretend objects and pretend actions (see section 3.5).

Learning Actions From Non-expert Demonstrations

The agent's model of affordance-based action variant generation was learned from repeated iterations of batch learning from non-expert demonstrations as a data set collected from five non-expert improvisers pantomiming pretend actions using the provided props as pretend objects. Non-expert improvisers were used as a data source to better reflect the target population for the improvisational experience created using this improvisational agent (for more, see section 1.3.2). After the demonstration sessions were completed, the data were processed by research staff using a separate annotation tool (see image 3.7) along with the help of recorded videos of the improvisational data collection session. After annotating the pretend action and pretend object as well as segmenting the start and end of the actions in the data collection session using between five and seven annotators, 893 mimed actions of length from 3.3 seconds (minimum length chosen) to 10 seconds (maximum length chosen) were curated as the initial data set. Each training data point (each action) was represented using the gesture vector and semantic vector presented in section 3.3.4. This data set has increased over time, but for a better comparison of results, later experiments were conducted using the same initial data set.

3.3.5 Improvisational Response Strategies (RQ2)

Improvisational response strategies extend prior work in LuminAI [50] and represent domain-independent procedural knowledge about the different reasoning strategies that human improvisers use during improvisation. While performing creative arc negotiation, these strategies enable human improvisers or improvisational agents to anchor their search for a suitable response to the current (or recent) improvisational context and reduce the size of their response search space and facilitate response generation in near real-time. The strategies themselves were compiled using literature search from jazz improvisation [58] and were adapted for use with dance [194] and improv theater [176].

The set of improvisational response strategies developed in LuminAI is extended in

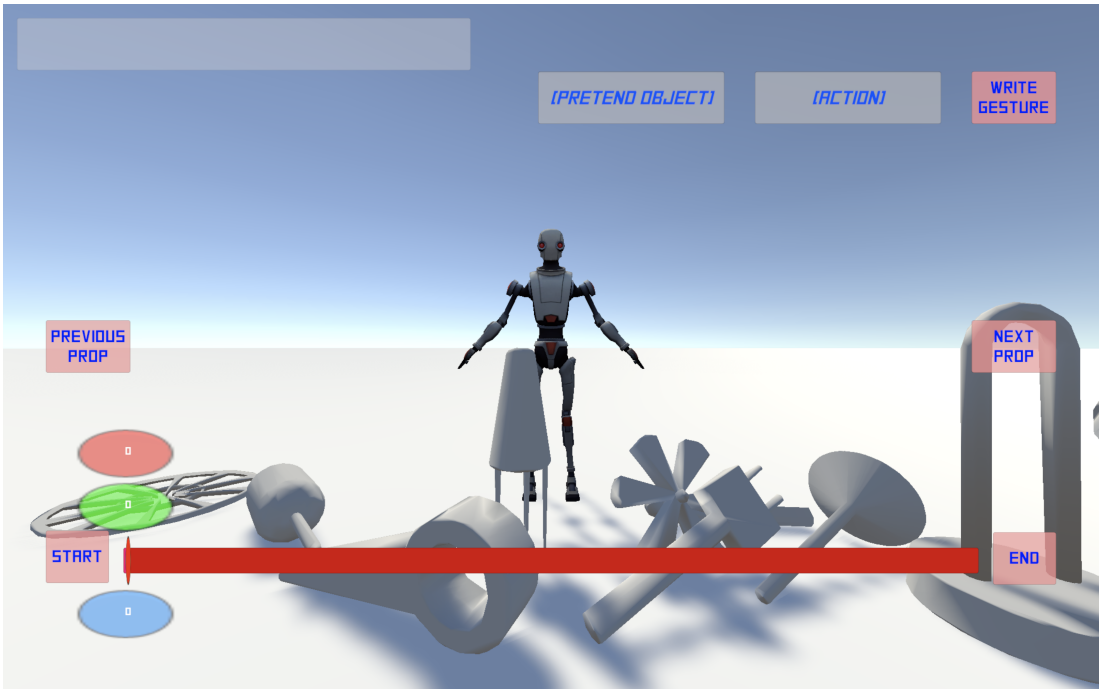

Figure 3.7: The annotation tool used to segment and annotate collected data.

this work. The strategies explored in this work include Mimicry, Transformation, Combination, Similarity-based Retrieval, and Pattern Projection. Other strategies for modulating the novelty, surprise, and value directly have also been proposed in initial ideation but remain future work. All of the strategies detailed in this research operate within the latent space of the CVAE generative model utilizing vector relationships that exist between points in that latent space. This provides a consistent mechanism for generating different action variants from a uniform representation and model. Some added detail for each strategy is given in the following paragraph, but specific implementation details in the CARNIVAL architecture are provided in the corresponding section of the architecture (see section 3.4.5).

Mimicry is the process of copying an observed action for interpretation and replay to for connecting to another improviser. *Transformation* consists of interpreting observed actions and then changing them according to relationships between previous actions. *Combination* involves interpolation of multiple recent actions as a way to generate variation while retain-

ing connections to prior context. *Similarity-based recall* can generate a space of gestures from most similar to least similar from the latent space. Finally, *pattern projection* is the process of using relationships between recent actions from the human and agent to project a source action to a target action so that it adheres to the given relationships.

3.3.6 Computational Models for Evaluating Creativity (RQ3)

The process of creative arc negotiation relies on the agent's ability to evaluate the creativity of perceived actions (and generated action candidates) computationally so that the agent can select responses that follow a target creative arc over the course of the improvised performance. Therefore this research contributes a set of computational models for evaluating creativity. The specific working definition of creativity used in this work follows Boden's definition of artifact creativity stating that creative actions perceived or generated by the agent are characterized by their degree of novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value). See section 3.4.4 for more information.

Framework For Creativity Evaluation Models

The models used in this research can be analyzed and potentially separated from other related computational models of creativity evaluation from the literature using a general framework. The framework specifies several dimensions along which the agent's approach to creativity evaluation are specialized due to the specific constraints of movement improv between a human and virtual character (as well as a potential audience). The general framework developed in this work is listed below.

1. The *perspective* being evaluated: There are three separate perspectives for judging the creativity of improvised interactions for an improvised performance/interaction between a virtual character, human collaborator, and an audience. The choice would depend on the main goal of the interaction, whether to optimize the quality of agent's learning and data acquisition, the user experience of the human collaborator, au-

dience enjoyment, or some combination of these (ideal for an improvised human-computer performance).

2. The degree of *dynamism*: This is the amount that the evaluation changes over time due to the experiences of the agent. A static/unchanging model would be fixed (without accounting for habituation or other changes over time), while a more dynamic model might adapt offline in between every improvisational session. The most desirable model adapts online over the course of the ongoing improvisational session.

3. The role of *feedback*: The model may not use feedback at all to improve its scoring over time. Alternatively, the model might utilize explicit feedback from the audience (e.g., applause) or collaborator (e.g., post-interaction surveys). The feedback could also be implicit through metrics like interaction duration or facial expression counts if explicit feedback can't easily be collected. Feedback is usually desirable unless the expertise of the system is far greater than the user.

4. The relative *expertise* of the system: A fledgling system that has little data or experience cannot expect to match human ratings of novelty and expectation and should treat the user's experiences as a superset of its own (e.g., an open-ended narrative improv system). A system that has collected data over its lifetime or through massive datasets can potentially surpass the human in terms of experience (e.g., a recipe generation system mining from large online recipe databases). It might then need to localize novelty and surprise estimation to the neighborhood of the user's experiences.

5. The relative *domain-dependence* of the model: Individual components of models for evaluating creativity can be considered on a spectrum from domain-independent to domain-dependent. For example, a theoretical model for evaluating novelty that uses aggregated distance measures between percepts in a given perceptual space can be considered largely domain-independent since the model could be applied to any

domain where percepts can be compared in perceptual space. On the other hand, a model for evaluating the quality (as a measure of value) would need to be more domain-dependent due to the high specificity of quality heuristics to a domain. In practice, all models lie on a spectrum somewhere between the two extremes, since some domain-specific knowledge is needed to operationalize the former and some domain-general processes can be used to apply the latter across domains.

Evaluating Novelty

Novelty of a perceived or generated action is evaluated in the agent using a distance-based comparison to other comparable actions that the agent has encountered before or is aware of (adapted from [141]). The distance-based comparisons are performed on both the gestural and semantic components of actions (see Section 3.4.4 for more detail about novelty calculation). Since the selection of all comparable actions to a specified action in the general case is a difficult problem, a naive solution is to compare against all actions perceived or generated by the agent. However, due to the growth of the agent's experience over its lifetime, this problem is approximated by comparing the specified action against its K nearest neighbors with K set empirically. The use of K nearest neighbors approximates the problem since, for truly novel actions, even the K nearest neighbors would be distant, while for commonly experienced actions, the K nearest neighbors would be at a short aggregate distance.

Evaluating Surprise

The agent computes the *unexpectedness* of a certain action as a measure of the *surprise* experienced by the agent. There are two general classes of methods that are used for computing unexpectedness in the related literature (see section 2.4). *Impact-based surprise* computes the impact of an observation on the agent's beliefs or expectations. *Deviation from expectation* is another class of models for computing surprise where distance-based

methods are used to compare the observed action and the most expected action(s) for a given situation. Both types of surprise are calculated in this agent to provide a balance between the two approaches. As discussed above in the framework for comparing creativity evaluation models (see section 3.3.6), this particular evaluation model for unexpectedness retains an exclusively agent-centric perspective. Future work could add additional perspectives or use personalization to tailor the perspective of the model over time to the participant or audience.

The distributions of expectation used in the unexpectedness evaluation models are conditioned on the physical attributes of the object (or prop) given to the agent (or human) due to the lack of data about the probabilities of pairs of actions over time in this particular domain as well as the feasibility of collecting this data. Future work could use interactive learning to approximate this distribution over time as the agent explored sequences with more of the actions in its action space. This would be an important step for adding narrative or causal coherence to the improvisation in order to apply this to future domains with added causal structure like full-scale embodied narrative improvisation.

Evaluating Value

Value is strongly dependent on the context, culture, and domain of the evaluation being performed and is a concept that is complex and overloaded in its use (see section 1.1 for definitions). Therefore, the agent evaluates the *quality* of perceived or generated actions as a measure of value using a set of heuristics specified for the Props game domain (it ignores all explicit reasoning about the societal or cultural value in its judgments of quality). Quality is more domain-dependent than novelty or unexpectedness (see discussion in section 3.3.6) and requires domain-dependent heuristics for its calculation.

A characteristic problem with improvisational domains like movement improv, in general, or the Props game, in particular, is the lack of a well-defined goal space or easily-specified objective function(s). Therefore, the heuristics defined for the agent in the Props

game domain are considered weak heuristics that are necessary but not sufficient and can thus only evaluate an incomplete region of the goal-space for the domain. Currently, in the agent, the two heuristics used to measure the quality of perceived or generated actions are the smoothness of a performed action, and the recognizability of a performed action in terms of the pretend object and pretend action that is signified by it (see section 3.4.4 for more detail). These heuristics compute the quality of an action both in terms of its gestural quality (though recognizability evaluates the gestural component in relation to the semantic component of an action). The different components are equally weighted in the current iteration of the model but can have their relative weighting modulated according to empirical results or even personalized to user preference in the future.

3.4 CARNIVAL: Creative ARc Negotiating Improvisational Virtual Agent pLatform

The robot improviser character that plays the Props game with VR users in the Robot Improv Circus installation is controlled by the CARNIVAL (Creative ARc Negotiating Improvisational Virtual Agent pLatform) agent architecture. CARNIVAL uses creative arc negotiation to address the improvisational action selection problem for embodied agents that perform movement improv. The architecture consists of three high-level components — perception, reasoning, and action — that work together to enable the agent to improvise.

CARNIVAL's *perception* module receives VR tracking data of the human's gestures in the virtual environment in the form of user-segmented gestures and interprets it in terms of the real/fictional pretend object being portrayed, the pretend action being pantomimed with that pretend object, and its location in various dimensionality-reduced spaces. The *reasoning* module reasons in real-time about what action from its open-ended action space best fits the target location on the agent's creative arc, given the previous improvisational context up till that point and how much of the performance remains. The *action* module receives the generated action from the reasoning module and plays it back in the virtual

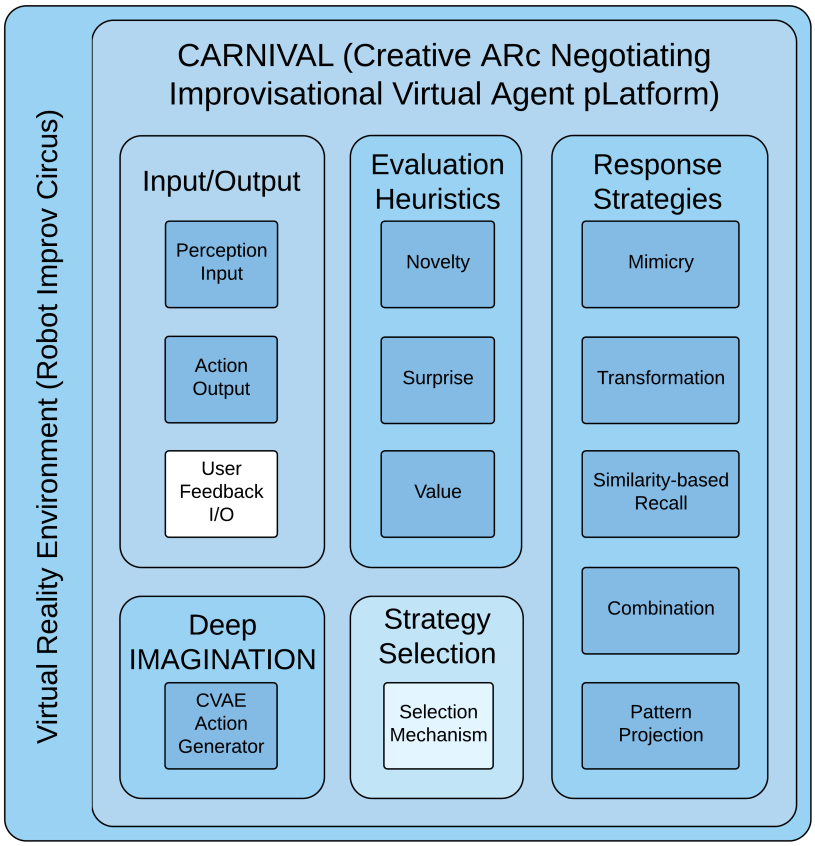

Figure 3.8: The CARNIVAL agent architecture that implements creative arc negotiation (see section 3.4.2 for process details). Lighter regions refer to future work.

world in a realistic manner by walking to the prop, picking up the prop, playing the generated action with that prop, dropping the prop when finished, walking to the buzzer, and hitting the buzzer to end its turn.

3.4.1 Perception: Interpreting Human Gestures

The perception module receives a gesture consisting of a temporal sequence of human pose data constituting a user-segmented gesture. Each frame of pose data is extrapolated from the instantaneous values of the three hardware-tracked points of a standard VR system (head, left hand, and right hand). The extrapolation is performed using inverse kinematics over the user's VR player avatar (the character used to represent the user in the virtual world). The end result of this process is world-space positions and rotations of the user's

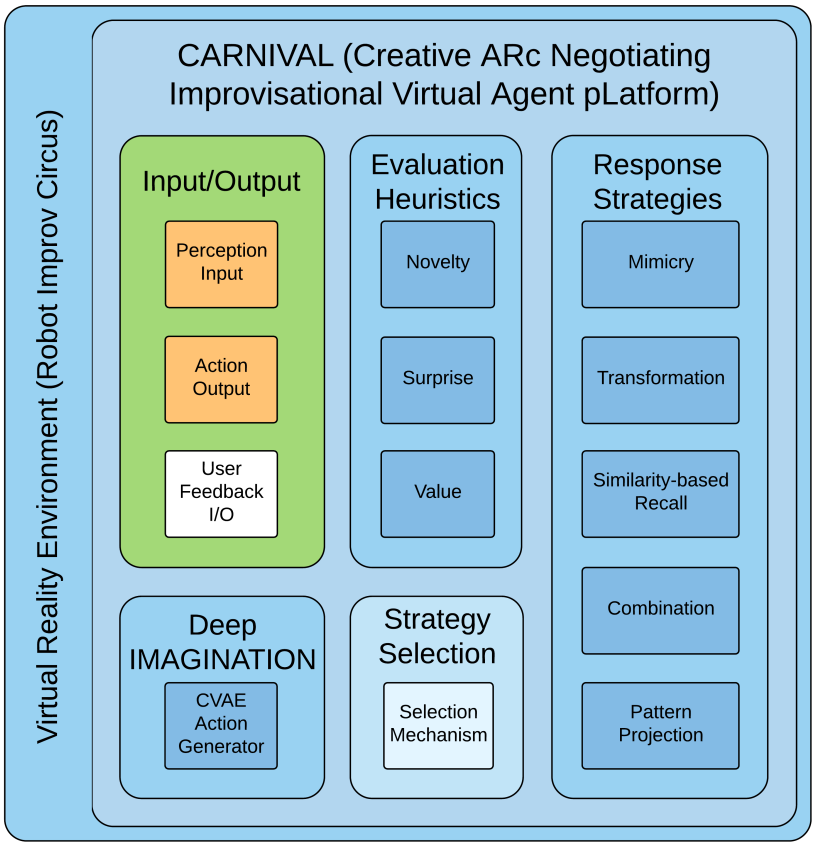

Figure 3.9: The CARNIVAL agent architecture with the perception module highlighted.

22 skeletal joints (head, neck, pelvis, left shoulder, right shoulder, etc.). In addition, the states of the VR controller buttons, as well as the position and rotation of the virtual prop on stage, are recorded in the perceived gesture.

Perceived gestures are vectorized into a 27000-dimensional or a 16000-dimensional vector representation. Both vector representations consist of concatenated frames of features extracted from the pose data. The 27000-dimensional vector uses 30 features per frame, extracted from the perceived gesture at 90 frames per second (FPS) for 10 seconds. The 30 features per frame consist of the normalized position (3D Cartesian coordinate representation) and normalized orientation (4D quaternion representation) of the user's head, left hand, right hand, and pelvis as well as the Boolean states of two VR controller buttons representing whether the user was trying to grab an object at that point. The 16000-

dimensional vector uses 35 features per frame, extracted from the perceived gesture at 45 FPS for 10 seconds. The 35 features per frame consist of the normalized position (3D Cartesian coordinate representation) and normalized orientation (4D quaternion representation) of the user's head, left hand, right hand, and pelvis, as well as the normalized position (3D Cartesian coordinate representation) and normalized orientation (4D quaternion representation) of the given prop. Zero padding values are added at the end for 250 places to round the total vector length to 16000. Additionally, if gestures are shorter or longer than 10 seconds, they are zero-padded at the end of the resulting vector or trimmed to the maximum duration, respectively.

Inferring Pretend Object and Pretend Action

Perceived gestures are interpreted by classifying them into pretend actions and pretend objects. The classification is done, at the moment, using a relatively simple K Nearest Neighbors classification approach. First, the dimensionality of the gesture vector is reduced to a two-dimensional point in the latent space of the generative model used to perform affordance-based action generation (see Section 3.4.3). The projected point is then used to query an RTree data structure [195] (a space partitioning tree with efficient dynamic loading that is widely used in spatial querying) for its nearest K neighbors (with K set empirically), consisting of previously seen and generated gestures. The RTree data structure is used as an optimization to perform K nearest neighbors search in logarithmic time complexity. The interpretation of a gesture in terms of an inferred pretend object and pretend action can be done using either a simple majority of its neighbors' object and action labels or by weighting the majority using their relative distances to the projected query point. The pretend action and pretend object are an English verb and an English noun (usually with high concreteness [196] respectively. The pretend action and pretend object are represented as 300-dimensional vectors from a pre-computed word embedding [197]. The respective vector interpretations of the gestural and semantic content of the action can

be dimensionally reduced using parametric T-SNE models [198] and are used later in the creativity evaluation process (see Section 3.4.4).

The RTree [195] that is used for nearest neighbor queries is initially pre-loaded with a data set of human-annotated actions [see Section 3.3.4] in order to avoid a cold start problem with the pretend action and object inference. These actions were annotated by members of the research team using a special annotation tool [See picture 3.7] while listening to participants talking through their improvisational performances with the given props. Since the RTree grows over the course of the installation as the agent's experience grows, it can be saved to a database for the agent's next run. The increase in elements also increases the size of K needed for the labeling process with the increase in RTree elements.

3.4.2 Reasoning: Creative Arc Negotiation

Interpreted human actions from the Perception module are received by the Reasoning module in order to generate the agent's response. The reasoning module uses creative arc negotiation to generate an appropriate response, addressing the improvisational action selection problem. This process is represented as an interruptible stochastic search through the creative space for a generated action that is nearest to the target point from the agent's creative arc for that turn or the agent's time remaining for the turn runs low.

Creative arc negotiation requires the following components to work together to enable the negotiation process.

1. The interruptible search process described above.

2. A parameterizable action variant generator that is able to search the agent's action space for candidate action variants.

3. A set of improvisational response strategies that can heuristically guide the agent's search to potentially important regions of the action space depending on the improvisational context till that point in the performance.

4. A set of computational models for localizing a generated action variant in the agent's creative space.

5. An optional process for selecting improvisational strategies according to the gradient between the agent's previous and current target points on the creative arc.

Creative Space and Creative Arc Representation

As described earlier (see Section 3.3.3), the agent's creative space is reductively adapted from Boden's definition of creativity as the novelty, surprise, and value experienced while evaluating a creative artifact. The agent's three-dimensional creative space consists of novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value), with each dimension having values measured in the closed interval [0.0, 1.0]. The creative arc is represented in CARNIVAL as a temporal sequence of creative space target points for the agent to use when selecting generated responses. The initial interpretation of a given creative arc is to treat it as a set of evenly spaced key points whose values can be linearly interpolated between based on the total number of turns for a given performance of the Props game.

Creative Arc Negotiation As Search

Creative arc negotiation is implemented as search through the agent's learned action space to find a generated action variant whose location in the agent's creative space is within an empirically set distance threshold ϵ of the current interpolated target point from the creative arc for the agent's turn. The search is interruptible so that if the agent's remaining time for the turn is equal to the maximum possible generated action variant length (currently 10 seconds), the nearest action till that point is returned as the agent's chosen response. The agent can perform creative arc negotiation either by considering the creative space locations of only its own actions or that of the human participant as well. In the latter case, the target point from the creative arc is compared with the halfway interpolated creative

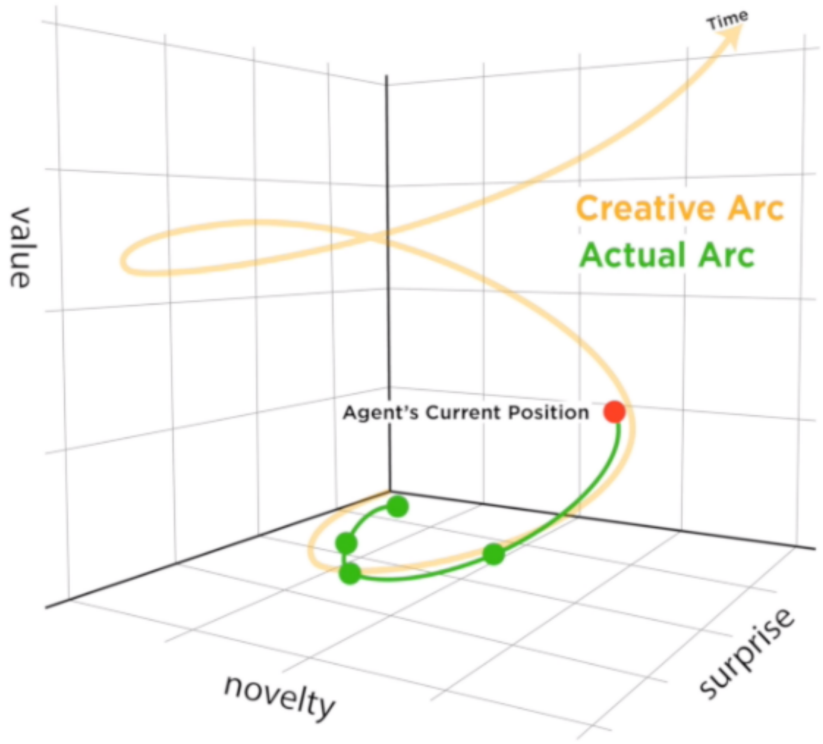

Figure 3.10: The actual negotiated creative arc in CARNIVAL.

space locations of the human participant's action and the agent's current generated action variant being considered.

3.4.3 Reasoning: Action Variant Generation

The process of searching through the agent's learned action space is operationalized using a parameterizable action variant generator. The action variant generator is trained on a data set of mimed human actions using props as real-world or fictional objects. A deep generative model is used to perform affordance-based action variant generation (see Section 3.4.3) by learning a parameterized mapping between the formalized physical attributes of the props used to mime the actions in the data set and the action space represented by the data set. The generative model's latent space is conditioned on the given prop's phys-

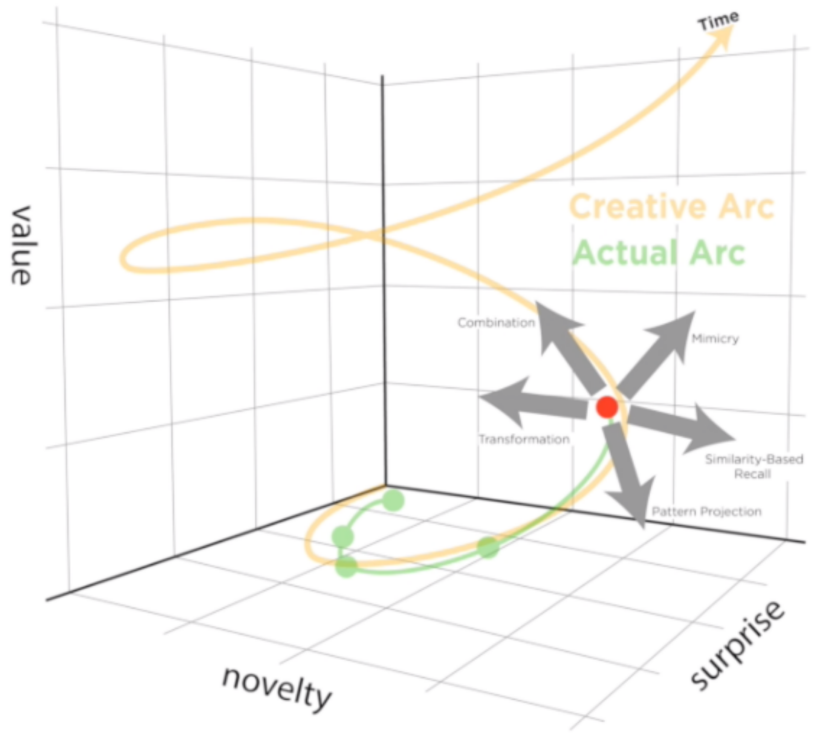

Figure 3.11: Improvisational response strategies used for guiding search through the agent's action space.

ical attributes and parameterized by a search control vector that can be varied to generate different actions mapped to the given prop's physical attributes. The iterative process of affordance-based action variant generation and evaluation by variation of the generative model's control vector is thus how the agent's action space can be searched in order to select responses to negotiate a creative arc.

DeepIMAGINATION

A deep generative model was used in CARNIVAL to perform affordance-based action variant generation, i.e., to learn the mapping between a prop's physical attributes and the set of actions that were shown to be possible to perform given props with those physical attributes.

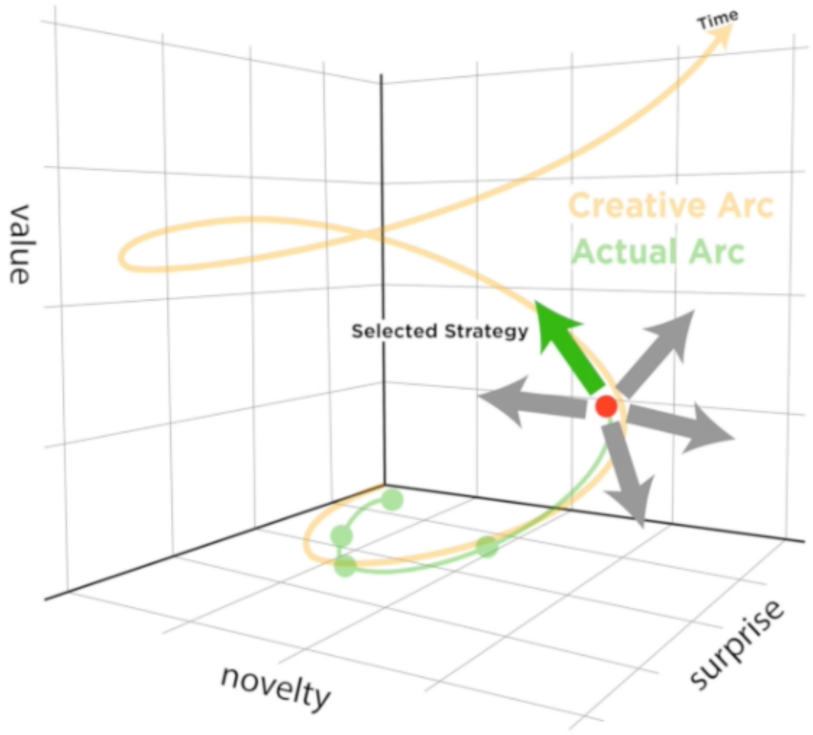

Figure 3.12: Optional strategy selection.

The conceptual model design was named DeepIMAGINATION for Deep IMprovised Action Generation through INteractive Affordance-based exploraTION [199] and was based around a general conditional variational autoencoder (CVAE) [185] architecture. Several variants of the actual model architecture were implemented using several convolutional and recurrent alternatives.

CVAEs consist of an encoder and decoder with conditioning happening on both the inputs to the encoder and decoder. In this case, the encoder and decoder were both conditioned on the physical attribute vectors of the props used to perform the actions using input concatenation. The encoder reduces the high-dimensional input into a low-dimensional latent space, and the decoder reconstructs a sampled latent vector back into a high-dimensional output from the same space as the input. Both convolutional neural

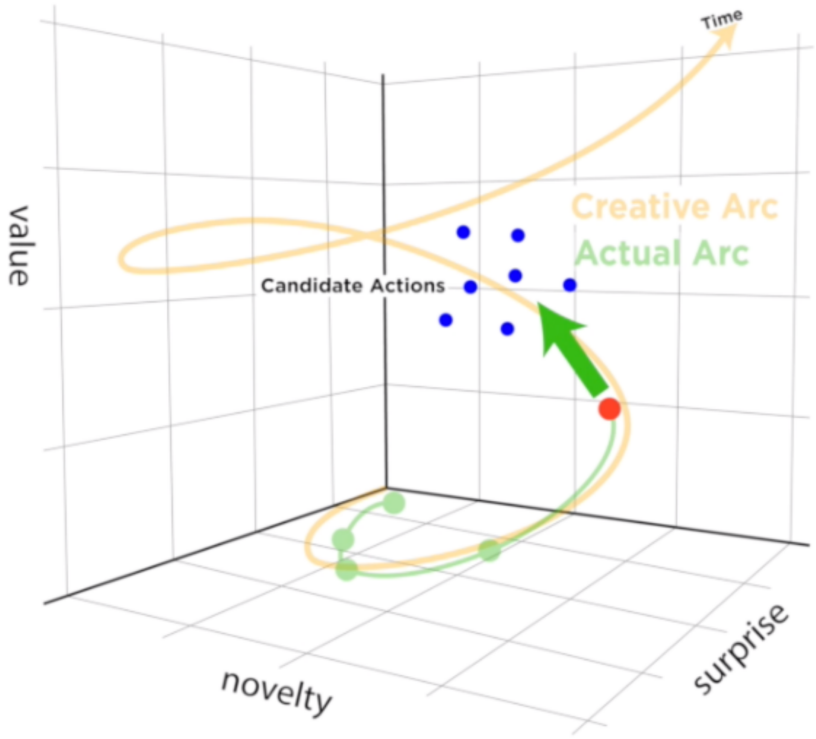

Figure 3.13: Local exploration of the agent's action space using DeepIMAGINATION.

networks (CNN) and recurrent neural networks (RNN) were used to implement variants of the DeepIMAGINATION module. One of the convolutional architecture variants is depicted in Figure 3.17. This particular variant uses 1-dimensional convolutional layers and 1-dimensional transposed convolutional layers in the encoder and decoder, respectively. Dropout layers were also used for regularization. A recurrent CVAE variant is described later.

Each CVAE variant was implemented in TensorFlow [200] and trained with the ADAM optimizer [201]. Each CVAE variant was trained on 900+ mimed pretend actions of length ranging from 3.3 to 10 seconds collected from five novice improvisers playing the Props game within a VR data collection environment (see previous Section 3.3.4 for details of the data collection). Therefore given an input distribution X, a latent distribution z and a

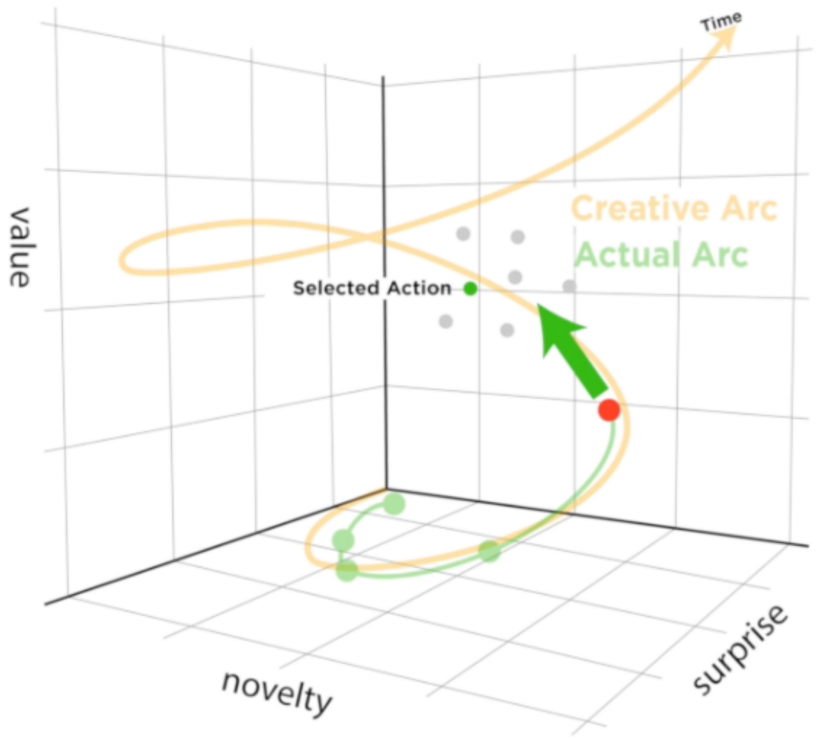

Figure 3.14: Action selection from the explored set.

conditioning distribution c, each CVAE variant was trained using the CVAE loss function defined as:

$$L(X, z, c) = E[\log P(X|z, c)] + D_{KL}[Q(z|X, c) \mid\mid P(z|c)] \tag{3.1}$$

In other words, the loss function is the sum of the decoder's reconstruction loss and the encoder's Kullback-Leibler divergence [202] loss, both conditioned on the physical attributes distribution. Training the network is made possible by using the re-parameterization trick (with mean $\mu(X, c)$ and diagonal covariance matrix $\Sigma(X, c)$) [187]:

$$z = \mu(X, c) + \Sigma^{\frac{1}{2}}(X, c)\,\epsilon, \text{ where } \epsilon \sim \mathcal{N}(0,\, 1) \tag{3.2}$$

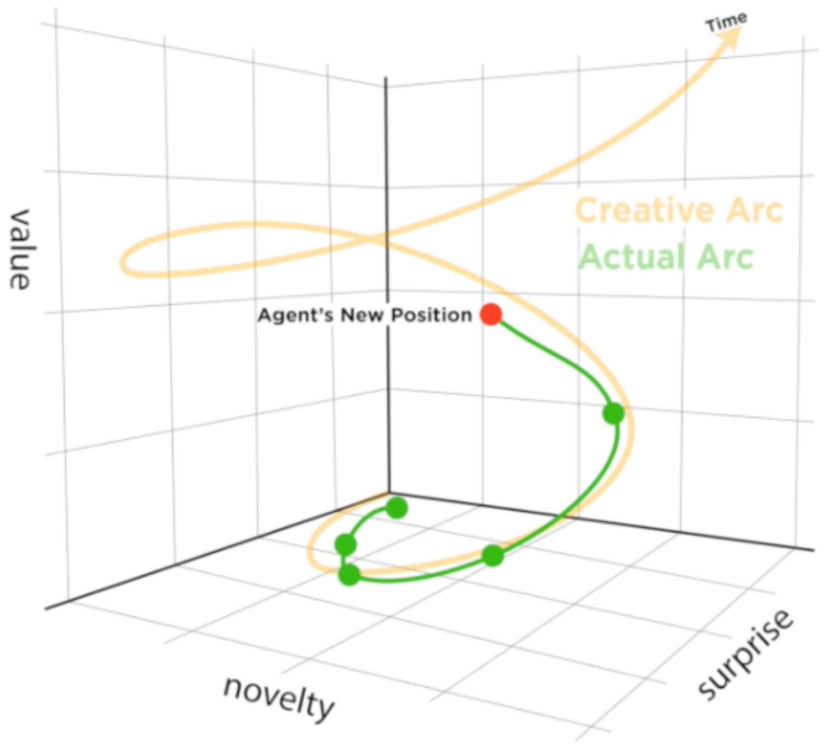

Figure 3.15: The negotiated creative arc in the agent is updated.

During the generation of action variants, the model's latent space is repeatedly sampled at specific locations provided by CARNIVAL's improvisational response strategies (see Section 3.4.5), based on the current improvisational context occurring. The DeepIMAGINATION module generates action variants conditioned on the physical attributes of the given prop. Generated action variants are evaluated by CARNIVAL's creativity evaluation models (see Section 3.4.4).

A total of eight architecture variants were designed and trained, including four convolutional models and four recurrent models. The variants were implemented for performance evaluation and selection (see evaluation experiments in Section 3.5). The convolutional architectures only differed in their input vector representations and resultant layer dimensionality. The four recurrent models used either a vanilla RNN architecture or an architecture

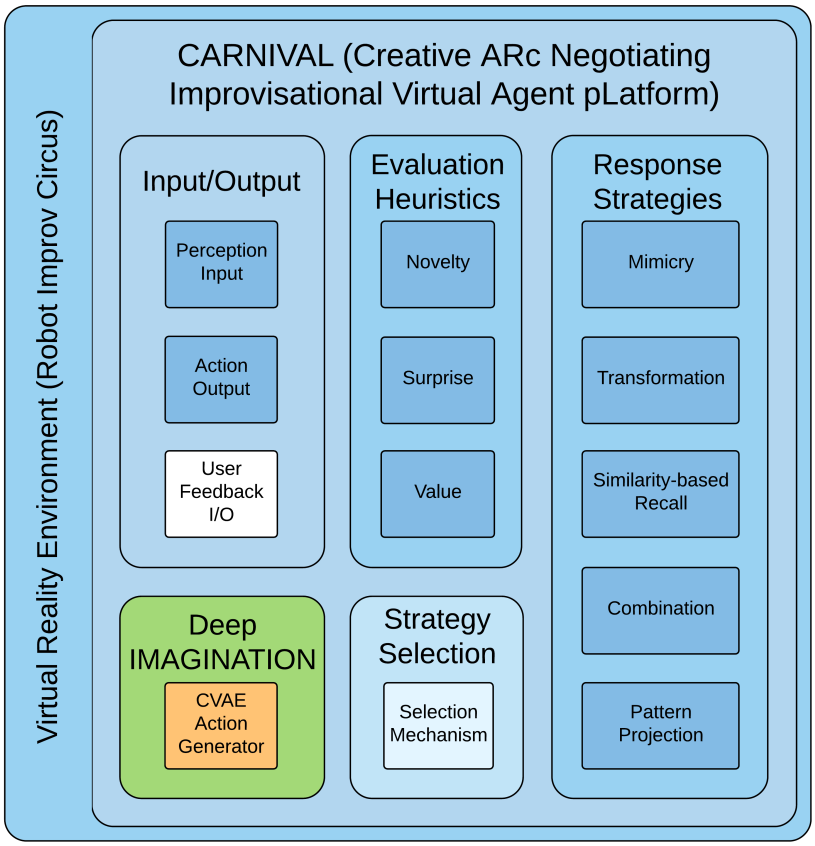

Figure 3.16: The CARNIVAL agent architecture with the DeepIMAGINATION module highlighted.

based on the MusicVAE network [203]. The two groups of RNN variants were also trained on different input vector representations.

Convolutional Variants

It is helpful to think of the different input vector representations ($(27000, 1)$, $(16000, 1)$, $(900, 30)$, and $(450, 35)$) for convolutional models in terms of the number of channels in the input data. The data was first represented with one channel, that is, 27000 and 16000 dimensional vectors were reshaped to $(27000, 1)$ and $(16000, 1)$ dimensional tensors, respectively. In another representation, the number of channels corresponded to the number of features per body pose frame - i.e., 27000 dimensional vectors were reshaped to $(900, 30)$ tensors while the 16000 dimensional vectors were reshaped to $(450, 35)$ tensors (discard-

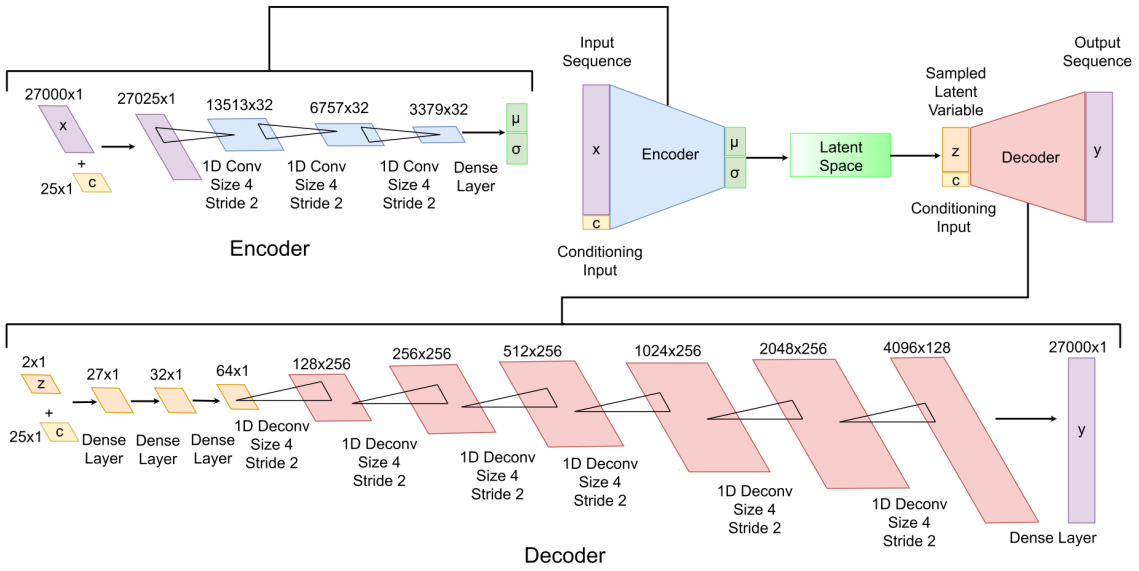

Figure 3.17: A convolutional variant of the DeepIMAGINATION architecture with $(27000, 1)$ shaped input gesture and 2D latent space. General CVAE architecture shown in upper right quadrant. Zoomed-in views of encoder and decoder in upper left and bottom respectively. Dropout layers not shown but applied between each convolution layer and between each transposed convolution layer.

ing the zero-padding). The outputs from the decoder were 27000 and 16000 dimensional vectors depending on the input vector representation.

Recurrent Variants

The RNN versions of CVAE were implemented using Long Short-Term Memory (LSTM) layers. Both the encoders and decoders of the Vanilla RNN implementation include single layers of bidirectional LSTMs that represented information for each frame concatenated with the physical attributes vector. Based on results from Roberts, Engel, Raffel, Hawthorne, and Eck, where vanilla RNN-based decoders sometimes had poor sampling and reconstruction performance, a hierarchical RNN architecture for the decoder was designed based on their MusicVAE architecture Roberts, Engel, Raffel, Hawthorne, and Eck. In this variant, the latent vector z is first passed through a fully connected layer to initialize the state of the Conductor layer, which is composed of a unidirectional LSTM layer. The output of the conductor layer is then passed as initialization for the bottom LSTM layers,

where each frame vector from Conductor layer, concatenated with the output of previous bottom layer LSTM, is used as initialization for the bottom layer LSTM of next time interval. The outputs of each bottom layer LSTM are then concatenated and flattened to match the input tensor shape.

3.4.4 Reasoning: Computationally Evaluating Creativity

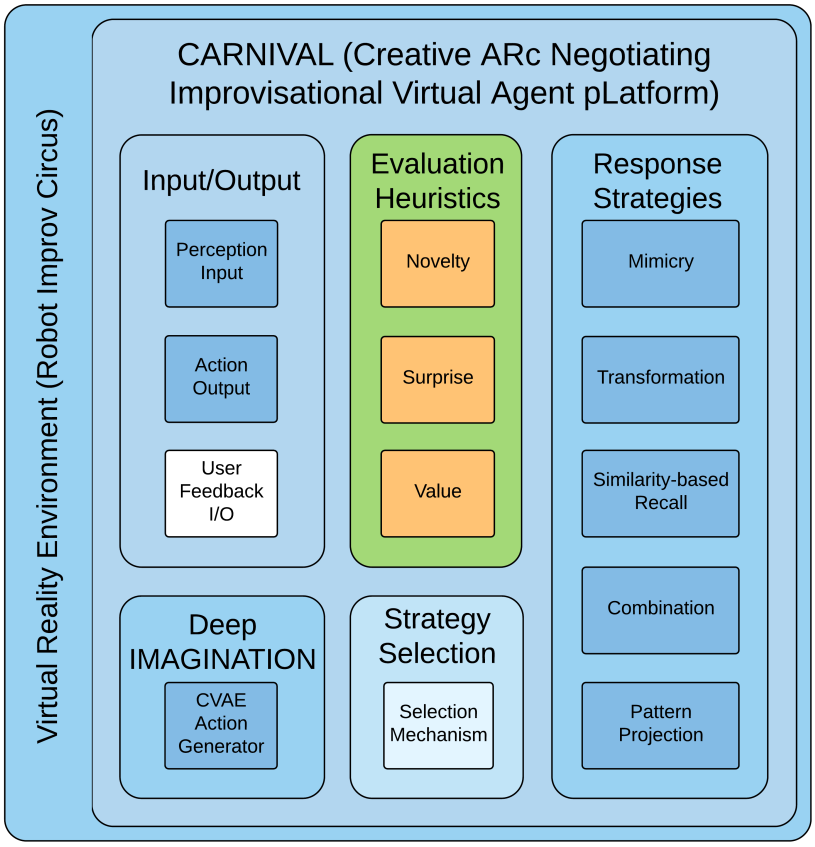

Figure 3.18: The CARNIVAL agent architecture with the computational models for evaluating creativity highlighted.

Perceived human actions and generated action variants from DeepIMAGINATION are localized to the agent's creative space and compared to find the response that is the local best fit with the current target point on the agent's creative arc. This process involves computationally evaluating the creativity of the perceived or generated actions. As described previously, the working definition of creativity in this work is the degree to which

a perceived or generated action is novel, surprising, and of high value from the evaluating agent's perspective. Therefore, the task of computationally evaluating creativity is divided into the subtasks of evaluating the novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value) of a perceived or generated action variant from the agent's perspective.

The operationalized definition of creativity in this work uses unexpectedness and quality as (imperfect) measures or proxies for surprise and value. Unexpectedness is used in the definition of creativity instead of surprise because the requirements for surprise as it is defined in this work specify it to 1) be an affective reaction and 2) cause a reaction proportional to the confidence with which a violated expectation is held [204]. As a result, the function over unexpectedness that is evaluated in the improvisational agent (see section 3.4.4) cannot technically be called surprise because it is not generated through any validated computational model of affect yet (such as one based in appraisal theory [205] or the somatic marker hypothesis [206]) and because the evaluation does not yet formally threshold the evaluated unexpectedness according to the measured confidence of the expectation. Both these shortcomings of the unexpectedness evaluation model will be addressed in future work.

The agent measures the quality of perceived and generated actions. However, the measured quality cannot necessarily be considered equivalent to value due to 1) the complex nature of defining quality metrics for the open-ended, ill-defined domains studied in this work, 2) the complex nature of 'usefulness' when dealing with the societal context of interactive installations or performative improvisation, and 3) the often complex functional relationships between the quality and value [207, 140]. These shortcomings of the work are planned to be addressed in future iterations of this work by studying 1) how people relate to this work as participants and observers, 2) practice-based and observational measures of quality within this creative domain, and 3) how it fits within the societal context of improv theater at multiple levels of analysis in the wild.

Evaluating Novelty

Novelty has been used to describe how new, unique, or original a sensation, percept, or experience is to an agent [208, 163, 209]. Through this definition, sensory percepts that an agent experiences can be evaluated for their novelty. However, generated items that arise from an agent's cognition can also be evaluated similarly as though perceptually experienced by the agent.

Novelty is measured in this research as the aggregated difference between a percept and other comparable experiences that an agent has already experienced (see 1). Using the framework discussed in section 3.3.6 about desirable properties of creativity evaluation models, this definition has the following properties. It is dynamic, since the experienced novelty changes over time, depending on the agent's experiences. It is evaluated from the perspective of the agent since the comparisons are made based on the agent's experiences. It does not incorporate external feedback to the agent to tune the model with human interaction. It is also designed to work most efficiently when the agent is relatively inexperienced compared to its human collaborators. Finally, the model is largely domain-independent though it does need domain-specific knowledge for both converting percepts into a spatial representation for applying distance-based comparisons and aggregating component novelty scores together to give a total novelty score. Future work would be needed to incorporate human feedback about the novelty evaluations of the model to improve its believability and to change the model to function more efficiently as it gains expertise from interacting with many human collaborators over time.

Novelty evaluation in CARNIVAL is modeled using the following algorithm.

It should be noted that finding the K nearest neighbors is used in algorithm 1 as an initial solution to the problem of selecting the set of comparable elements against which to compare the percept. The value of K was set empirically by using the elbow method [210]. This method was used on a graph of the absolute acceleration of the mean of mean distances between every K nearest neighbors in a collected data set of size N for values of

Input: Percept X, Integer K, and RTree<PerceptVector> R
Result: The general novelty score for an observed percept X

PerceptVector XVector := DimensionalityReduce(X);
PerceptVector[] NearestKNeighbors := FindKNearestNeighbors(XVector, K, R);
Distance[] NearestKDistances := GetDistances(XVector, NearestKNeighbors);
Double Novelty := Mean(NearestKDistances);
R := UpdateRTree(XVector);
Return Novelty;

Algorithm 1: ComputeNovelty(. . .)

K ranging from 2 to $\frac{N}{2}$.

CARNIVAL's novelty model evaluates perceived and generated action variants. Since an action is composed of gestural content and semantic content (in the form of the pretend action and pretend object), novelty values are calculated for each of the three components, aggregated together, and scaled to the closed interval [0.0, 1.0]. The exact process for each type of novelty proceeds as follows (see 1).

Gestural content in an action is represented as either a 27000-dimensional or 16000-dimensional vector. Since nearest neighbor searches are conducted over the gesture vectors to find its K nearest neighbors, the high-dimensional vectors representing gestural content are dimensionality-reduced using parametric T-SNE [198]. Parametric T-SNE is a manifold-learning approach to non-linear dimensionality reduction implemented using a neural network encoder trained on a parametric T-SNE loss function. The parametric T-SNE model is advantageous over newer T-SNE dimensionality-reduction implementations that are faster such as Barnes-Hut T-SNE [211] since the model [198] does not have to be repeatedly retrained to work on new data, though the learned transformations would be increasingly distorted over time without retraining.

The K nearest neighbors of the dimensionality-reduced gesture vector are found using an RTree data structure [195] that performs the query in logarithmic time complexity. The mean distance between the evaluated percept and its K nearest neighbors is used as the gestural novelty component score. This score component is aggregated with the semantic

novelty component score and scaled to get the final novelty score.

Semantic content in an action consists of two English words representing the action's pretend action and pretend object, for example, 'looking through' and 'kaleidoscope'). These are represented as two 300-dimensional word vectors from a pre-computed word embedding [197]. The dimensionality of each word vector is reduced using a similar PT-SNE model as the gestural content but trained on an English word data set. The dimensionality reduced output vectors are used to query two separate RTree data structures [195] that are populated with previously experienced pretend actions and pretend objects for the respective sets of K nearest neighbors. The respective mean distances are calculated from these sets of neighbors. These pretend action and pretend object novelty score components are averaged to get the semantic novelty score components.

The total novelty score is calculated by computing the mean of gestural and semantic novelty component scores. The resulting total score is adaptively scaled to get a final score in the closed interval [0.0, 1.0]. Adaptive scaling is performed so that the expected minimum and maximum values of the source domain can be adjusted according to the observed minimum and maximum values generated by the creativity evaluation models.

Evaluating Unexpectedness

The creativity evaluation models in CARNIVAL measure the unexpectedness of a perceived or generated action variant as a proxy for the surprise that an agent might encounter in these situations. Unexpectedness is defined in this research in terms of the degree that an experience deviates from the agent's expectation for that experience. The degree of surprise is also proportional to the confidence of the agent's belief or expectation, i.e., the higher the agent's confidence in a belief or expectation, the higher the agent's surprise if it is violated [212].

Surprise differs from novelty in subtle but significant ways. Novelty is a global measure of the difference between an experience being evaluated and other comparable experiences,

regardless of what an agent might expect that experience to be like. In contrast, surprise is the difference between an experience being evaluated and what it expects that experience to be, regardless of how different that experience is to other comparable experiences. For example, let us imagine a hypothetical situation where an agent is shown different images of animals from the African savanna and is asked to identify the animal. In that context, if the agent is shown an animal that it has seen examples of many times, say a giraffe, it can correctly identify the animal as a giraffe. The animal is not novel to the agent. Additionally, given that the agent is expecting to see images of animals from the African savanna, the image of the giraffe is not surprising to it. The agent is then shown an image of a newly discovered animal called an 'eleppo' which looks like a combination of an elephant and a hippopotamus. The eleppo is different from all the other animals that are potentially comparable to the eleppo and is thus novel to the agent. The eleppo is also quite different from all the animals the agent was expecting to see in an African savanna context, therefore, the eleppo is surprising to the agent.

The pattern of showing the hypothetical agent a familiar animal and then an unfamiliar animal created from combining other animals is repeated several times, continuing the example. The agent has built up expectations about many aspects of the game by this time, including the categories of animals it is shown each turn, the shifting context of each round of the game, and specifically that the second round of each pair will see an animal combined from two other animals in the African savanna. At this point in the game, the combination animals the agent is shown in every even-numbered turn is still different from every other comparable animal it has seen before and is thus novel to the agent. However, the agent has correctly learned to expect that it will see an animal that is made up of a combination of other animals found in the African savanna and is thus not surprised by the novel animal when it sees an image of that animal.

Continuing the example further, in the next even-numbered turn, however, the agent sees a familiar giraffe again. The image of the giraffe violates the agent's expectation of

seeing an unfamiliar combination animal like the eleppo. Therefore, the agent is surprised at seeing the familiar image of the giraffe. The image of the giraffe is not novel to the agent though since it has seen many images of giraffes like that one in the past.

The preceding example clearly demonstrates two scenarios where novelty and surprise can differ significantly. However, differences between surprise and novelty are often not as clear cut. Additionally, it is not always the case that all expectations that are violated are based on temporal patterns. In the Props game that is played within the Robot Improv Circus installation, unexpectedness can be generated from atemporal sources of expectation violation. In that case, props that are given to the player or agent can convey expectations for their usage as certain pretend objects over others. For example, a large prop that consists of two spherical parts joined together by a long thin cylindrical part, would be more commonly expected to be used as a cartoon barbell rather than a comically large swizzle stick. These expectations are atemporal in nature depending on the implementation of the agent's models, i.e., not necessarily dependent on the temporal ordering of actions observed by the agent. However, since the degree of unexpectedness is proportional to the confidence with which a belief is held or expectation is generated, beliefs that are reinforced or weakened over time may give rise to more or less unexpectedness if violated, as in the case of an agent that has a dynamic model of unexpectedness implemented. For a detailed taxonomy of dimensions along which to inspect the various kinds of expectation used for evaluating surprise, see [213].

In the literature, approaches to measuring surprise or unexpectedness have been divided into those that measure the impact of an experience on an agent's prior beliefs and those that directly measure the deviation of an observed experience from the expected outcome or experience [213]. This research uses both methods to compute an aggregated score for the unexpectedness of a perceived or generated action variant.

The agent's model for evaluating unexpectedness can be analyzed using the framework introduced in section 3.3.6. The agent's model of unexpectedness is dynamic since it is

updated according to the agent's changing beliefs. Therefore, the agent would be less surprised over time if it repeatedly observed the initially unexpected percept in a given context. The model measures unexpectedness from its own perspective rather than that of the human collaborator or a potential audience. While the model does update over time through belief updation, it does not yet tune its outputs according to external feedback in order to potentially make its evaluations more realistic. Additionally, the model is currently designed to be more efficient for situations where the agent has less expertise than a human collaborator or audience, such as an interactive learning context where the agent improves over time. The agent's model of unexpectedness is also largely domain-independent and could be applied to other domains in a straightforward manner with appropriate knowledge about what distributions of beliefs would be relevant for the model in the new domain.

The agent's general model for evaluating unexpectedness uses a combination of Bayesian Surprise [59] and direct computation of 'deviation from expectation' [141] (DFE). This model can be seen in algorithm 2. The model computes unexpectedness over the gestural and semantic content of a perceived or generated action variant. The two scores are then adaptively scaled to the closed interval [0.0, 1.0], as with the novelty score components, and the mean is returned as the total unexpectedness score for the agent.

Input: Percept X, Integer K_{BS}, Integer K_{DFE}, and RTree<PerceptVector> R
Data: ProbabilityDistribution PerceptDistribution
Result: The unexpectedness score for a given percept X computed from Bayesian Surprise and deviation from expectation methods

```
PerceptVector XVector := DimensionallyReduce(X);
Double BSScore := ComputeBSScore(XVector, K_BS, R, PerceptDistribution);
Double ScaledBSScore := Scale(BSScore, 0.0, 1.0);
Double DFEScore := ComputeDFEScore(XVector, K_DFE, PerceptDistribution);
Double ScaledDFEScore := Scale(DFEScore, 0.0, 1.0);
Double Unexpectedness := Mean(ScaledBSScore, ScaledDFEScore);
R := UpdateRTree(XVector);
PerceptDistribution := UpdateDistribution(XVector);
Return Unexpectedness;
```

Algorithm 2: ComputeUnexpectedness(. . .)

The component of the agent's unexpectedness that is computed based on the impact that an experience has on the agent's prior beliefs is based on Bayesian Surprise [59] and can be seen in algorithm 3. Bayesian Surprise is the aggregated difference between a probability distribution of the agent's beliefs before (prior distribution) and after (posterior distribution) observing some percept, and the difference between the prior and posterior distributions is calculated using KL Divergence [202]. According to this measure of surprise, the higher the change on a prior belief due to observing some evidence, the higher the surprise. Bayesian Surprise is computed across the gestural and semantic content of an action variant separately and summed together.

Input: PerceptVector XVector, Integer K_{BS}, RTree<PerceptVector> R, and ProbabilityDistribution PerceptDistribution
Result: The unexpectedness score for a given percept X computed using the Bayesian Surprise method

```
ProbabilityDistribution Prior := GetPriorDistribution(PerceptDistribution);
ProbabilityDistribution Temp := Clone(PerceptDistribution);
Temp := UpdateDistribution(XVector);
PerceptVector[] NearestKNeighbors := FindKNearestNeighbors(XVector, K_BS, R);
Temp := UpdateDistibution(NearestKNeighbors);
ProbabilityDistribution Posterior := GetPriorDistribution(Temp);
Double BSScore := ComputeKLDivergence(Prior, Posterior);
Return BSScore;
```

Algorithm 3: ComputeBSScore(. . .)

The other component of unexpectedness computed in the agent's model is a direct measure of the degree to which the gestural and semantic content of a perceived or generated action variant differs from the most expected set of action variants. The algorithm for calculating this, in general, is shown in algorithm 4. The process is repeated for the gestural and semantic components of the action variant, and the resulting scores are summed together.

Many different probability distributions could have been used to compute the Bayesian Surprise and DFE components of unexpectedness. However, at least initially, the agent's expectations are conditioned on the given prop rather than on other temporal distributions (like the temporal expectation of which pretend action could follow the last one in

Input: PerceptVector XVector, Integer K_{DFE}, and ProbabilityDistribution PerceptDistribution
Result: The unexpectedness score for a given percept X computed using the deviation from expectation method

```
PerceptVector[] MostExpectedKPerceptVectors :=
  FindKMostExpectedPercepts(PerceptDistribution, K_DFE);
Double DFEScore := GetMeanDistance(XVector, MostExpectedKPerceptVectors);
Return DFEScore;
```

Algorithm 4: ComputeDFEScore(. . .)

a narrative). Specifically, the generated expectations are based on the conditional probability distributions — P(dimensionality-reduced pretend action vector|prop physical attributes), P(dimensionality-reduced pretend object vector|prop physical attributes), and P(dimensionality-reduced gesture vector|prop physical attributes). In the future, other conditional probability distributions could also be useful for calculating the unexpectedness component scores. These could include probability distributions such as P(dimensionality-reduced pretend action at time=t_n|dimensionality-reduced pretend action at time=t_{n-1}) leading to more temporally or causally coherent action variants. However, these distributions are not currently used, due to a lack of data at present.

Evaluating Quality

The agent evaluates the quality of a perceived or generated action variant using a set of domain-dependent heuristic functions as a simplistic measure of its value. At present, this set of heuristics consists of the *smoothness* of the gestural content of an action variant and the *recognizability* of the gestural content of an action variant (recognizability from the agent's perspective). The current two heuristics were chosen by considering the aesthetics of a performed gesture and by consulting domain experts in improv theatre for their suggestions as well. They could be expanded in the future to include measures of coherence or even humor (given a computational model of humor). The general algorithm combining the smoothness and recognizability components can be seen in algorithm 5.

Input: Action A, Integer K, and RTree<GestureVector> R
Result: The quality score for a given action A computed from smoothness and recognizability heuristic functions

```
GestureVector GVector := DimensionallyReduceGesture(A);
SemanticVector[] SVectors := DimensionallyReduceSemantic(A);
Double SScore := ComputeSmoothnessScore(GVector);
Double ScaledSScore := Scale(SScore, 0.0, 1.0);
Double RScore := ComputeRecognizabilityScore(GVector, SVectors, K, R);
Double ScaledRScore := Scale(RScore, 0.0, 1.0);
Double Quality := Mean(ScaledBSScore, ScaledDFEScore);
R := UpdateRTree(GVector);
Return Quality;
```

Algorithm 5: ComputeQuality(. . .)

The agent's model of quality evaluation can be analyzed using the framework stated in section 3.3.6. The quality evaluation model is less dynamic than the novelty and unexpectedness evaluation models since the smoothness component measures the same intrinsic property of all gestures that it evaluates. However, the smoothness component can't be used to filter out gestures that score low in that heuristic since the agent might require a lower smoothness gesture during improvisation or the recognizability might boost quality to a required level according to the current creative arc. In contrast, the recognizability heuristic is dynamic because its estimation of recognizability would change over time based on the relative frequencies of observed gestures for each semantic class (classes of pretend actions and pretend objects) that a gesture could be labeled with as well as their relative distances to each other. The agent evaluates quality purely from its own perspective rather than from that of a human collaborator or audience. The quality evaluation model does not incorporate external feedback at the moment, though future versions could feasibly use a trained classifier to recognize aesthetic quality based on feedback from collaborators or audience members. Like the previous models described in this work, the agent's models of quality evaluation are also based on an agent with less experience than its human collaborators or audience members. Unlike the previous two models, the agent's quality evaluation model is heavily domain-dependent. This is because quality itself, unlike novelty and unexpect-

edness, is heavily tied to the conventions, rules, and boundaries of the specific domain in which it is evaluated.

The smoothness of an action variant is shown in algorithm 6. The smoothness heuristic function measures the average jerk [214] in the motion of each joint in a gesture across three different windows sizes for aggregating the motion. This results in three vectors of movement at different scales of resolution per joint of the agent's pose frame. In physics, jerk is computed as the third derivative of a positional vector (first two being velocity and acceleration). Therefore smoothness is computed as the inverse of the mean jerk across all joints for an agent across three resolutions of movement. An inverse scaling is used in the heuristic since high jerk equates to low smoothness.

Input: GestureVector GVector
Data: Integer LocalWindowSize, Integer RegionalWindowSize, Integer GlobalWindowSize
Result: The quality score for a given gesture vector GVector computed from the smoothness of GVector

```
DoubleVector[] JointVectors := GetJointVectors(XVector);
Integer[] WindowSizes := [LocalWindowSize, RegionalWindowSize,
 GlobalWindowSize];
DoubleVector MeanJointJerkValues;
for Integer WindowSize in WindowSizes do
    DoubleVector[] AvgPoolJointVectors := AvgPoolVectors(JointVectors,
     WindowSize);
    DoubleVector JointJerkValues := ComputePerJointJerk(AvgPoolJointVectors);
    MeanJointJerkValues := IncrementalMean(JointJerkValues, 1);
end
Double SScore := Mean(MeanJointJerkValues);
Double ScaledSScore := InverseScaling(SScore);
Return ScaledSScore;
```

Algorithm 6: ComputeSmoothnessScore(. . .)

The recognizability of an action variant is shown in algorithm 7. Recognizability is intended to calculate the degree to which a gesture can be recognized as a specific pretend action or pretend object. This is calculated by finding the K nearest neighbors to the observed gesture. From this set, the mean distance between the observed gesture and all

gestures with pretend action and pretend object interpretations matching the observed gesture (intra-class mean distance) are found. The mean distance between the observed gesture and neighbors with pretend action and pretend object interpretations that don't match the observed gesture (inter-class mean distance) are found next. The ratio between the intra-class and inter-class mean distances is inverse scaled and returned as the recognizability score.

It may not immediately be obvious how the scaled inverse of the intra-class to inter-class mean distance ratio predicts recognizability. Examining the heuristic, this function scores low when the inter-class mean distance is low, and intra-class mean distance is high. This implies that the nearest gestures that do not match the observed gestures' semantic interpretation are nearer to the observed gesture than the nearest gestures that do match the observed gesture's semantic interpretation. The opposite holds for a high score of this heuristic, i.e., previously observed gestures that match the observed gesture's semantic interpretation are nearer to it than previously observed gestures that do not. This implies that in the former case, the gesture does not 'look like' other gestures that have been interpreted the same way, and in the latter case, it does.

The models for evaluating novelty, unexpectedness, and quality return scores in the closed interval [0.0, 1.0]. The scores form a three-dimensional point in the agent's evaluative creative space. The scores are used by the creative arc negotiation process to find the closest generated action variant to the next target point on the creative arc.

3.4.5 Reasoning: Improvisational Response Strategies

Creative arc negotiation could be performed with an exhaustive search of generated action variants in the creative space. However, given that a primary characteristic of improvisation is the necessity to return satisfactory responses in near real-time with a potential loss of optimality (if optimality were even possible in movement improv), the creative arc negotiation is necessarily guided by heuristics to seek out potentially lucrative regions of the

Input: GestureVector GVector, SemanticVector[] SVectors, Integer K, and RTree<GestureVector> R
Result: The quality score for a given gesture vector GVector computed from a recognizability heuristic

```
GestureVector[] NearestKNeighbors := FindKNearestNeighbors(GVector, K, R);
SemanticVector[] NearestKLabels := LabelGestures(NearestKNeighbors);
GestureVector[] LabelMatchGestures := FindLabelMatches(NearestKNeighbors,
  NearestKLabels, SVectors);
GestureVector[] LabelMismatchGestures :=
  FindLabelMismatches(NearestKNeighbors, NearestKLabels, SVectors);
Distance[] MatchDistances := GetDistances(GVector, LabelMatchGestures);
Double MeanMatchDistance := Mean(MatchDistances);
Distance[] MismatchDistances := GetDistances(GVector, LabelMismatchGestures);
Double MeanMismatchDistance := Mean(MismatchDistances);
Double RScore := MeanMatchDistance ÷ MeanMismatchDistance;
Double ScaledRScore := InverseScaling(RScore);
Return ScaledRScore;
```

Algorithm 7: ComputeRecognizabilityScore(. . .)

search space as quickly as possible. These heuristics are encoded in CARNIVAL through improvisational response strategies.

Improvisational response strategies were introduced earlier in this chapter as formally encoded strategies used by improvisers to generate near real-time responses during improvisation by bounding their search to the current improvisational context. Searching the agent's action space in CARNIVAL is performed by varying the parameters to the DeepIMAGINATION module. Therefore, improvisational response strategies are encoded as strategies for generating parameters for DeepIMAGINATION based on the current improvisational context. The generated action variants are then evaluated using the creative space localization models.

Mimicry

Mimicry is a response strategy where the agent observes the human's action, interprets it in terms of its learned action space, and attempts to generate the action it just saw the human perform. In contrast to prior work in the LuminAI agent [50], where the agent

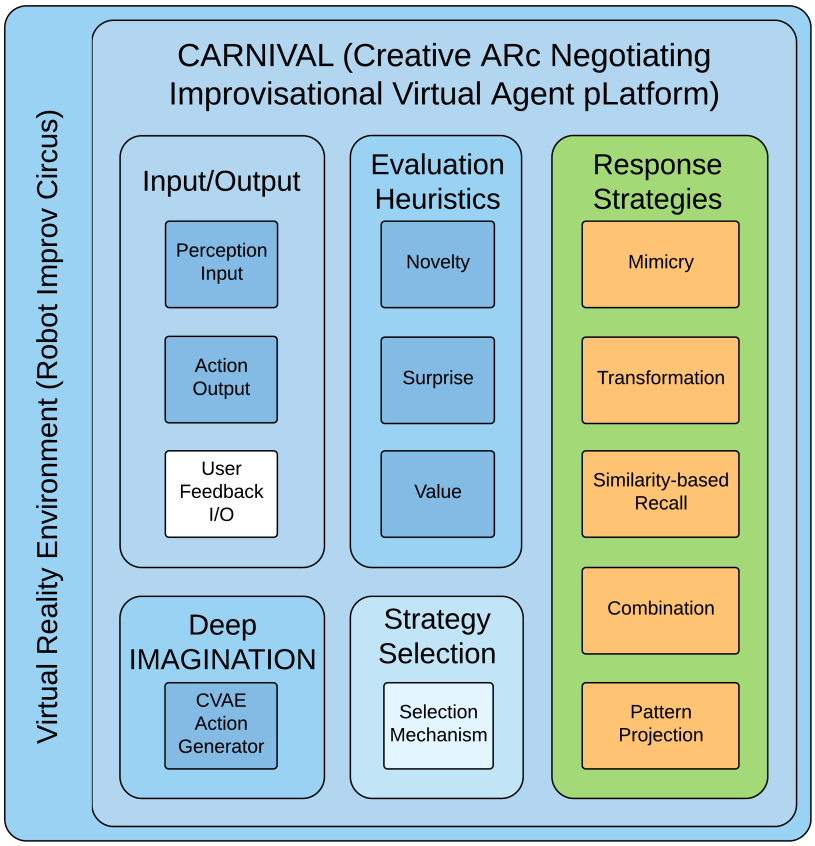

Figure 3.19: The CARNIVAL agent architecture with the improvisational response strategies highlighted.

could playback the human's gesture exactly, CARNIVAL agents map the observed action to a point in its learned action space (i.e., as a point in DeepIMAGINATION's latent space), and then try to recreate or regenerate it using DeepIMAGINATION. Therefore, the agent will be able to mimic the user's action to a high degree of accuracy only if it has been trained on actions similar to what the human just performed. The converse of that statement is also true, and the action variant CARNIVAL generates through mimicry may not look exactly like the human action depending on its novelty to the agent. This is more realistic in terms of what a human might expect from another human improviser in terms of both process and result since it is unlikely that a human would be able to perfectly recreate an action that is largely novel to them on their first try as well. The lack of exact replay is further justifiable, given that CARNIVAL is designed to be retrained regularly on obtaining more data from

human participants over time.

Combination

Combination is a response strategy where the agent can interpolate between N actions that are similar (but not the same) to the current action from the agent's past experience and combine them together to create a new action variant. When $N = 1$, this is a purely episodic recall response strategy, i.e., being reminded of a similar action from the agent's recent experience and no combination is actually done since there is only 1 action. However, when $N > 1$, this is a unique strategy where the agent is reminded of N actions, and they are combined together into a novel action variant. When $N = 2$, the strategy performs an interpolation in latent space similar to other generative models [203, 215]. Combination for $N > 2$ is achieved by finding the centroid of the coordinates of the N similar (but not the same) actions in DeepIMAGINATION's latent space and then generating the resulting combined action variant.

Transformation and Pattern Projection

Transformation as a response strategy was previously used in prior work as changes that could be made to a gesture according to the specific aspects of the gesture and other functional changes in the gesture's form itself. In CARNIVAL, transformations are performed by doing vector operations between coordinates in the latent space being used to generate actions from DeepIMAGINATION.

The use of the DeepIMAGINATION latent space as a representation of the agent's learned action space is convenient for implementing action space search as well as for the application of response strategies as search control. However, the latent space in DeepIMAGINATION is not directly interpretable, i.e., each dimension does not correspond to meaningfully higher-level or usefully abstracted dimensions. Therefore, it is currently difficult to map interpretable semantics onto the latent space dimensions in order to directly

generate transformations based on them like in prior work [50]. Transformations are thus based on patterns that are calculated between previous human-agent turns.

Pattern projection can directly be applied by finding a vector between the actions in the previous human and agent turns in DeepIMAGINATION's latent space and then translating that vector to the coordinate of the human's current action in the latent space. This applies the properties of the spatial relationship between the two actions in the previous turn to the current human action. Baseline pattern application without any other transformations added to that pattern is thus translation in the vector space. Patterns can also be found by looking at the current action perceived by the agent, searching for the closest action from the agent's episodic memory (temporally backwards), and then looking at how that action was responded to from the episodic memory, calculating the vector relationship between those two actions and projecting that vector out from the current action in DeepIMAGINATION's latent space.

Additional affine transformations can be added to a translated vector in the latent space. These can include rotation, reflection, and magnitude scaling of the pattern or spatial relationship vector. Reflection is interpreted as applying a complementary (if not opposite) pattern to the current action. Further research is required to better understand the interpretations of spatial relationships or patterns in the latent space and to find interpretable dimensions that can be mapped onto the latent space.

Similarity-based Recall

A simplified and constrained version of *Similarity-based recall* was previously presented in prior work [50]. In prior work, this strategy was restricted to finding the most similar gesture in an interpreted space. CARNIVAL's version of the similarity-based recall response strategy is a significant advancement over the previous version of this strategy due to the parameterizable nature of similarity in this search. The agent can recreate the most similar (but not same) recent action as its response, equivalent to how this was implemented in

prior work. However, it can also recreate the least similar (or most dissimilar) recent action. It can even generate an action variant that is arbitrarily in between the two extremes of most similar and most dissimilar. Currently, the most similar and most dissimilar strategies are used.

Generating Novelty, Unexpectedness, and Quality

Previous experience with the strategies in the LuminAI architecture [2] had qualitatively indicated some trends in how people perceived the generated gestures from the different response strategies. Mimicry seemed to be perceived as low novelty, but participants described the experience of seeing the character repeat their actions very positively (high value from the user's perspective). Transformations tended to be perceived with varying degrees of quality but had high estimations of novelty. The moment when characters did not repeat a user's actions but did something different was also rated as highly unexpected. Similarity-based Recall and Pattern Application in LuminAI were perceived similarly to Mimicry (lower novelty and higher quality), though they also created moments of unexpectedness in participants when they realized that the agent was not only performing mimicry. These trends from prior work in LuminAI need to be examined in the future and determined whether they transfer to the CARNIVAL architecture and the Robot Improv Circus installation as well. Strategies unique to or approached differently in CARNIVAL, such as combination, pattern application, and similarity-based recall are currently not known in terms of their predicted effects on participants for evoking novelty, unexpectedness, and quality. Other strategies for CARNIVAL were also proposed in [176] for modulating the perceived novelty, surprise, and value directly and could be added in the future.

Strategy Selection

Strategy selection was originally meant to be a part of CARNIVAL as a way to further optimize the creative arc negotiation process. It was originally meant to be implemented

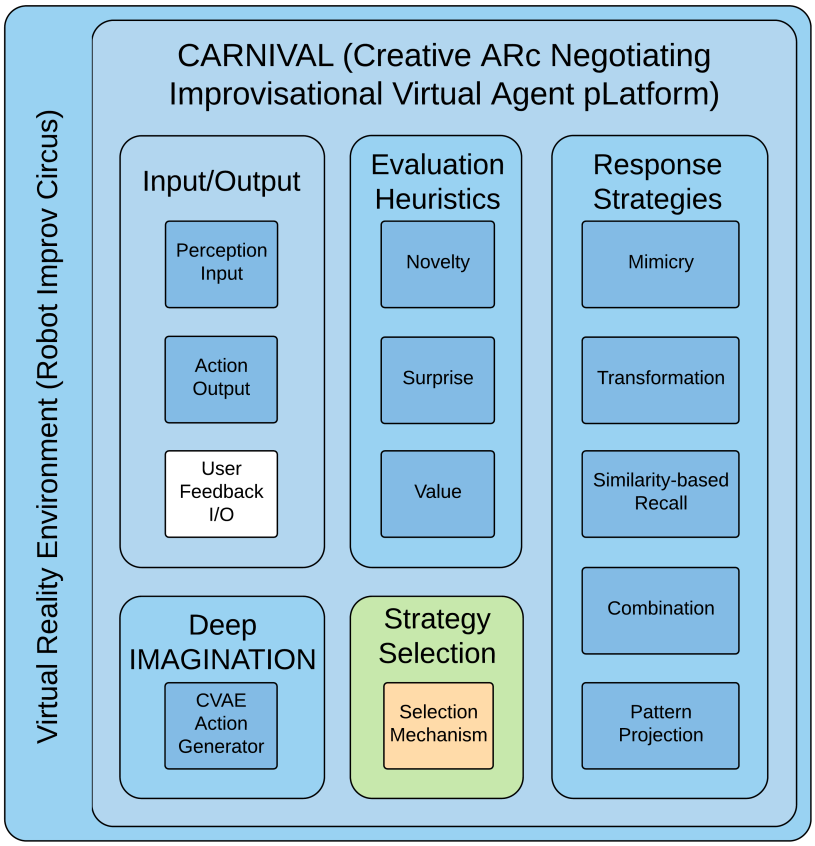

Figure 3.20: The CARNIVAL agent architecture with its naive implementation of strategy selection as parallel strategy execution highlighted.

using a learned policy mapping between the known set of strategies and the desired direction of movement in the agent's creative space between the agent's current location and the next target point on the creative arc. However, for the initial iteration of the CARNIVAL architecture, it was decided that all strategies would be executed in parallel and strategy selection would be applied in the future.

3.4.6 Action: Performing Agent Responses

The action module in CARNIVAL receives a selected action variant once the creative arc negotiation process is completed. The action module uses a finite state machine, inverse kinematics (IK), and path planning over a navigation mesh (or navmesh) to perform the action variant more realistically on stage. These features are implemented using off the

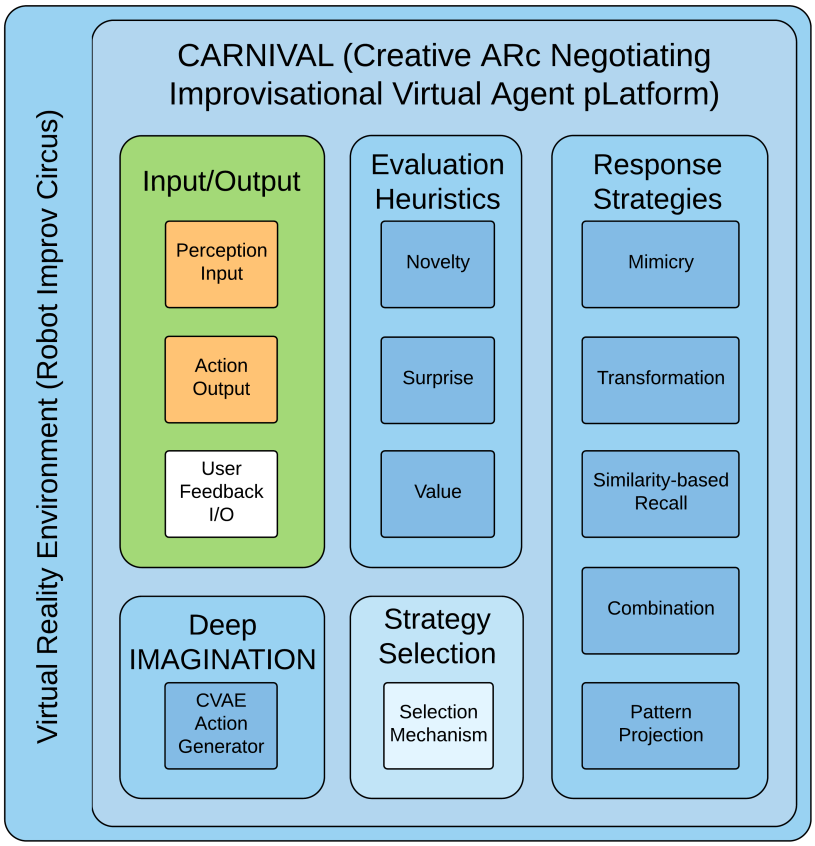

Figure 3.21: The CARNIVAL agent architecture with the action module highlighted.

shelf tools or included features within the Unity3D game engine.

Realistic Action Playback

The action module is implemented as a finite state machine (FSM) that receives actions and plays it back. When it is the agent's turn in the round/game, and the reasoning module outputs an action variant for performance, the action FSM first transitions from an idle state to a walk-to-prop state. The agent uses path-planning over a navmesh to walk to the current location of the agent's prop for that turn. When at that location, the FSM then transitions to a pick-up-prop state where it uses IK to move its hand down to the prop on the ground, attach the prop to its hand, and stand back up using IK again. The FSM then transitions to a walk-to-first-location state, and the agent navigates back to the location on stage where the

first frame of the action variant will start. The agent then plays back the action variant using IK, with the prop attached to the correct hand or controlled by the gesture representation directly. After finishing playback of the selected action variant, the FSM transitions to a walk-to-stage-center state and the agent walks back to the center of the stage. Then the FSM transitions to a drop-prop state, and the agent drops the prop by detaching it from its hand. The FSM then transitions to a walk-to-buzzer state, and the agent navigates to its buzzer. The FSM then transitions to a hit-buzzer state, and the agent uses IK to hit the buzzer. This ends the agent's turn, and the FSM switches back to an idle state. For the duration of the human's turn, the agent turns to watch the human's performance.

3.5 Evaluation

The research presented thus far in the chapter described the Robot Improv Circus installation and the CARNIVAL architecture as tangible boundary objects for studying the claims made in my thesis statement. My thesis statement stated, "embodied agents that address the improvisational action selection problem using 'creative arc negotiation' increase perceptions of enjoyment, agent creativity, and coherence in both observers and participants while performing movement improv with non-experts." My research questions served as a guiding outline for evaluating the claims made in my thesis statement over the course of this research and are repeated for convenience as follows.

RQ1 How can an agent perform parameterized action variant generation from a learned action space based on the physical attributes of a given object?

RQ2 How can an agent improvisationally search its action space based on previous experience and the current improvisational context?

RQ3 How can an improvisational agent computationally evaluate the creativity of perceived or generated actions in near real-time in terms of their novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value)?

RQ4 How can an embodied agent select actions to negotiate a given creative arc in order to address the improvisational action selection problem while performing movement improv with non-experts?

RQ5 How does addressing the improvisational action selection problem while performing movement improv with non-experts affect both observer and participant perceptions of enjoyment, agent creativity, and coherence?

The following list describes the formal evaluation plan for the research questions repeated above in order to evaluate the claims made in my thesis statement systematically. I first describe two experiments to validate the implementations of affordance-based action variant generation (RQ1) and the creativity evaluation models (RQ3) in CARNIVAL. I then describe the next set of three experiments and aim to show that my thesis statement holds with the following chain of evidence that they provide.

1. The improvisational action selection problem is successfully addressed by creative arc negotiation as an approach.

2. Embodied agents addressing the improvisational action selection problem using creative arc negotiation can perform movement improv with non-experts.

3. Embodied agents addressing the improvisational action selection problem using creative arc negotiation and performing movement improv with non-experts can be interacted with as a participant and experienced as an audience member.

4. Embodied agents addressing the improvisational action selection problem using creative arc negotiation can perform movement improv with non-experts so that perceptions of enjoyment, agent creativity, and coherence increase for both participant and audience member.

3.5.1 Validating Affordance-based Action Variant Generation in CARNIVAL

DeepIMAGINATION is the parameterizable action generator for conditionally searching the agent's action space in the CARNIVAL architecture based on the physical attributes of objects given to the agent (see Section 3.4.3). It was designed with the aim of enabling an agent to search a learned action space in order to generate believable, recognizable, and high-quality pretend action variants with similar abstract props. Therefore, validating the component involved investigating whether the architecture allows an agent to generate action variants that were believable, recognizable, and high-quality compared to human actions? This was determined using a survey-driven study of non-experts evaluating human and computer-generated actions from DeepIMAGINATION in the criteria of believability, recognizability, and quality.

Methodology

Multiple surveys were created using Amazons Mechanical Turk platform that assessed the believability, quality, and recognizability of four data sets related to actions from DeepIMAGINATION (human actions, agent mimicry of human actions, near variants of mimicked human actions, and far variants of mimicked human actions). The experiment was conducted to address the evaluation question described above. Each of the four data sets consisted of 40 gestures performed by a robot character in VR across 20 props from the Robot Improv Circus. A GIF was recorded of the robot character performing two actions with each prop for a total of 160 actions across all four datasets. These GIFs were then evaluated by remote workers on the Mechanical Turk platform.

The human-generated data set comprised actions performed by a human in VR with a robot avatar. This set of human gestures was then passed through DeepIMAGINATION in various conditions to generate three additional data sets of actions with the same robot avatar. The direct output of the autoencoding process made up the agent mimicry data set as it represented the agent's interpretation of human gestures. The third and fourth data

sets were made up of near and far action variants (respectively) of the agent mimicry data set. They were generated by sampling points at a radius of 0.1 and 2.0 away (respectively) from the mimicry gestures in the CVAE model's latent space with the exact values for radial distance determined empirically. The same robot avatar performed these actions as well.

Each survey required the participant to watch either one or two recorded GIFs of actions (depending on the task involved) from one of the four datasets and answer questions about the GIF(s). In each survey, the human data set made up the human actions, and the other three data sets made up the computer-generated actions. There were 80 participants for tasks with single GIFs (absolute ratings) and 60 participants for tasks with two GIF comparisons (comparative rating in a forced-choice configuration). Each participant worked on 20 GIFs out of the entire data set of GIFs.

Believability: In order to assess the believability of the actions, two survey tasks were given to Mechanical Turk workers. In the first survey, each participant watched a single GIF (absolute rating configuration) at a time and answered whether they believed the action was performed by a human in VR or generated by a computer program. The comparison was made in order to evaluate whether participants could tell the difference between computer-generated (CG) actions and human actions between each data set. The hypothesis was that differences would be seen between the discrimination accuracy of the generated actions according to which of the three CG data sets was being evaluated (indicating that at least some groups of CG actions were as believable as human actions). Notably, if the participant's accuracy at this task was low, it meant that the generated action was easily mistaken for human action, and thus, the generated action variants were believable.

A second task was conducted that asked people to compare a human action from the human actions data set with a CG action from one of the other three datasets and asked the participant to identify which action they believed was generated by a computer. The test helped to clarify whether participants thought that computer-generated actions were human actions when directly comparing the two. The test also indicated how believable

the CG GIFs were. If the participants had low accuracy in determining the identity of the CG GIF, it would indicate that the CG GIFs were believable. The hypothesis was that there would be significant differences in recognition accuracy across groups, indicating that the CG actions were mistaken for human actions in some of the groups.

Recognizability: The recognizability of the actions in the four data sets was assessed in terms of how accurately identifiable both the pretend object and pretend action were that the character in the GIF was portraying (no written annotations were given to them in the GIF, of course). The survey asked participants to select what they believed the robot character was most likely enacting from a list of three options. The options were similar to stabbing with a sword or eating with a spoon. High accuracy in identifying the actions and objects shown in the GIF would indicate that the portrayal was recognizable overall. Our hypothesis was that comparable recognition accuracy across groups would be seen showing that the CG action sets were equally recognizable to human actions.

Quality: Participants were asked to determine the quality of the GIFs through two tasks. In the first one, participants were asked to rate the smoothness and quality on a 5-point Likert scale by looking at a GIF and evaluating it on its own. They were also asked to state what their criteria were for quality in this domain before rating any GIFs and were asked to use those criteria strictly during rating.

The second task was designed as a comparative rating, forced-choice configuration task. Participants were asked which action of the two they thought was smoother and of higher quality. Each participant was asked to define quality themselves at the beginning of the survey and to strictly use those same criteria while rating the GIFs for quality later on.

The two measures (smoothness and user-defined quality) were used together to assess the overall quality of each action in both tasks. If smoothness and user-defined quality were high for each action, it would indicate that the overall quality was high. Our hypothesis was that there would be comparable quality and smoothness ratings across groups.

Results

Believability: The task of detecting whether a given GIF was human performed or CG was treated as a binary classification task between the performance of the participants on the human data set in comparison to their performance on each of the other three data sets. The lower the participant accuracy, the stronger would be the evidence that the CG actions were believable. In order to analyze the participant responses, a confusion matrix was created for the four sets of comparisons: human vs. all CG, human vs. agent mimicry, human vs. near variant, and human vs. far variant. The F1 scores for the four conditions were: 0.5251, 0.7154, 0.7163, and 0.671. Additionally, the Matthews Correlation Coefficients for the four conditions were: 0.3308, 0.4237, 0.426, and 0.2912, respectively (all weak positive correlations).

The results above showed that the believability of the CG actions was comparable to that of the human actions in the absolute rating task when human vs. all CG or human vs. far variant conditions were considered. The fact that far variants scored the highest in comparison to agent mimicry and near variants was surprising since it was the least close to the corresponding human point in the latent space. However, it possible that it was close to some other human point and thus ended up generating believable actions.

Responses from the comparative rating, forced-choice configuration study of believability between two action GIFs were assessed by treating the task as a multi-class classification problem. The options given to participants were – CG action on the left, CG action on the right, both CG actions, and neither CG actions. As a reminder, poor participant performance on this task would be indicative that the CG actions were highly believable.

A four-class confusion matrix was created for the four responses possible, once each for human vs. agent mimicry, human vs. near variant, and human vs. far variant. In that order, the F1 scores were 0.8157, 0.7925, and 0.7678, respectively. The Matthews Correlation Coefficient was calculated, respectively, to be 0.5414, 0.5938, and 0.4931 (strong positive correlations). Both results were calculated using micro-averaging due to the multi-

class condition. The result indicated that when compared directly side-by-side to a human-performed action, participants were able to identify the human-performed action with relatively high accuracy, indicating that the actions were not as believable as desirable when compared directly against a human-performed action.

Recognizability: Participants of the recognizability experiment were asked to identify the actions performed by robot characters when assessing the recognizability of actions. Their mean accuracy (standard deviation in parenthesis) was determined across the different data sets ordered as human, agent mimicry, near variant, and far variant as 0.64 (0.26), 0.37 (0.24), 0.41 (0.23), and 0.33 (0.27). The median accuracy values for the same groups were 0.66, 0.4, 0.4, and 0.30. This outcome is a negative result that shows that recognizability for CG actions was comparable to random guessing, while human-performed actions were twice as likely to be recognized correctly.

A Shapiro-Wilk [216] test found a non-normal distribution for the accuracy data. Therefore, a Kruskal-Wallis omnibus rank-sum test [217] was computed on the data. The results were found to be significant, and the null hypothesis was rejected with a p-value $= 5.505771 \cdot 10^{-17}$. A Dunns test adjusted with Benjamini-Hochberg FDR showed that all the negative result relationships between the human data and the CG data were significant (all p-values < 0.019641).

Quality: For the comparative rating, forced-choice configuration study of the smoothness and quality of each action, the medians were calculated for the Likert scale responses and chi-squared tests were calculated for the human data compared to each of the three data types to see if there were significant associations between the types of data and the Likert scale responses for smoothness (or high quality respectively). For absolute smoothness, the median values for human, agent mimicry, near variants, and far variants were 4, 2, 2, 3 on a 1 - 5 scale from not at all smooth to very smooth. The chi-squared test reported significance with $\tilde{\chi}^2$ = 304.9299 and a p-value < 0.00001. For absolute user-defined quality, the median scores reported for the same data sets were 4, 3, 3, 3 on a similar scale from

very poor quality to very high quality. The chi-squared test reported significance with $\tilde{\chi}^2$ = 265.4731 and a p-value < 0.00001.

When assessing the comparative rating, forced-choice configuration study of smoothness and quality for each action, the percentage of results that were considered smoother (or higher-quality respectively) was recorded along with chi-squared tests that were calculated for human data compared to each of the three data types. The test was conducted to see if there were significant associations between the types of data and the selection of the human or computer action as more smooth (or high quality respectively). For smoothness, human data was chosen as smoother 75.63% against agent mimicry, 77.54% against near variants, 75.30% against far variants, and 76.14% overall against all CG actions. There were no significant differences found between the groups, with $\tilde{\chi}^2$ = 0.6701 at a p-value < 0.05. For user-defined quality, the percentage of responses where human data was chosen as higher-quality was 73.58%, 78.26%, 76.74%, and 76.14% for the same ordering as smoothness. There was no significant association found either, with $\tilde{\chi}^2$ = 2.6957 at a p-value < 0.05.

Discussion

Survey-driven observer-ratings were used to validate the implementation of the affordance-based action generation module, DeepIMAGINATION, within the CARNIVAL architecture. The experiment was conducted by collecting observer-ratings of video clips of the agent performing different types of action variants in order to evaluate the believability, recognizability, and quality of the action variants with respect to each other. Ratings were performed either on each variant by itself or by directly comparing human and computer-generated variants together.

The survey-driven observer-rating study of believability, recognizability, and quality produced mixed results. The action variants generated showed high believability in the absolute rating configuration (single video clip rating) with users confused about whether it was a human or computer-generated clip roughly half the time. The believability of gener-

ated action variants in a comparative-rating configuration was less stellar with much higher accuracy for humans correctly identifying the two categories. This is understandable because the absolute rating case is a more realistic condition for evaluation since there would only be one action at a time in the actual Props game setting where this would be used. The comparative-rating configuration could then be considered a ceiling on performance for the agent's generation.

The results for recognizability were clearly negative. In the task, users were matching the pretend object and pretend action to three options each. The results for human actions were more than double that of generated action variants. Additionally, the rater's recognition accuracy for generated action variants was as low as random guessing performance. As a result of the low recognizability ratings from observers, the agent was given a speech bubble and an audible robotic voice (using text-to-speech) that announced what it was attempting to portray using template-based dialogue generation. An example can be seen in figure 3.3. A casual pilot interaction and experience design study with three participants was conducted with this feature both activated and deactivated, resulting in every participant corroborating the high utility of the speech bubble and audio voice for added participatory recognizability.

The quality of the generated action variants was evaluated in terms of smoothness and user-defined quality. These values were comparatively high for all types of evaluated data in the absolute rating configuration. However, there was a definite surprise in the comparative rating version of this task. In the comparative rating configuration for comparing relative quality between action GIFs (which was expected to be the performance ceiling condition), 25% of the time when comparing both smoothness and user-defined quality, raters preferred the generated action variant over the human action. This was a surprising result since it was expected that close to 0% of raters would choose the generated actions in this comparative condition.

There were certain methodological limitations to this study, as well. Firstly, the task of

evaluating generator outputs against each other (or by themselves), completely outside of any context, would have been quite unusual for many human evaluators without possessing a good reference point or comparison for what the expected bounds of performance were in this task (though perhaps less so for those familiar with prop-based improv theatre). Therefore, the first set of results of the human evaluation task may not have truly reflected the agent's performance within the context of the entire CARNIVAL architecture. Therefore, further studies have also been conducted to elaborate on the findings and limitations of this first study, culminating in observer rating and in-person evaluation of the entire CARNIVAL architecture as an improvisational partner.

3.5.2 Validating Creativity Evaluation Models in CARNIVAL

The agent's computational models for evaluating the creativity of perceived as well as generated actions in terms of their novelty, unexpectedness, and quality were studied through a set of validation experiments. The experiments were conducted using non-expert observer ratings through a set of survey-driven tasks. The aim was to show that the implementation of the computational models in the CARNIVAL architecture rated the perceived and generated actions similarly to non-expert human observers.

Methodology

An initial three-part study comparing the results of the agent's creativity evaluation models to the human perception of the novelty, unexpectedness, and quality of actions performed by an agent was conducted on the Amazon Mechanical Turk platform. The study was conducted with a sample of 50 online non-expert participants for each of the three properties being compared. At a high level, participants were provided with video clips to compare and asked to rate which video clip was more novel, more unexpected, or of higher quality.

Each participant was made to compare 20 pairs of video clips featuring the agent performing actions from the training data set (see section 3.3.4) originally collected for the

purposes of training DeepIMAGINATION (see section 3.4.3). Each pair was matched to either compare two actions with a high and a low rating (experimental condition) or compare two actions that both had medium ratings (control condition) for the property being studied (say novelty). The primary study hypotheses were that there would be statistically significant differences in the distributions of accuracy scores per participant for ratings that matched the evaluation model's choice for the higher-rated action between the experimental condition and the control condition in each of the three experiments involving novelty, unexpectedness, and quality, respectively.

Each of the tasks for novelty, unexpectedness, and quality were conducted as a separate task. Additionally, the novelty, unexpectedness, and quality ratings were calculated for all actions in our data set, and then for each task, different actions were chosen. In other words, the novelty evaluation task had different actions compared to the quality evaluation task (which differed from the chosen actions for unexpectedness in turn), since the low, medium, and high scoring actions would potentially be different for each axis of the creative space.

For each of the studies, the properties of actions that the users were rating were defined for the user before the task. However, there can be no guarantee that the definitions were solely used by the user to make their choice. Additionally, each user was asked to define for themselves criteria they would use to evaluate the creativity of an action performed with a prop in a movie, theatrical play, or session of pretend play. They were also asked to choose the video clip that featured the more creative action strictly using their previously-stated criteria. They were asked at the end why they picked any notable actions over others in terms of creativity as a way to get them to reflect on their decisions (even if it was a post hoc rationalization).

The definitions given to participants included the following concepts and definitions in language that tried to avoid being too technical. *Novelty* was defined as *how different or new or original the given action was to them in its performance as well as its intent*

compared to all the other comparable actions they might have been reminded of (even a little bit) while watching it. Object surprise was defined as *how unexpected the object was that the current prop was imagined (or pretended) to be, given the look and feel of the prop.* For example, according to that metric, given a long cylindrical prop, it might have been imagined to be a mop, which would be unsurprising; however, if it was imagined to be a limp spaghetti noodle, that would be surprising. *Action surprise* was defined as *how unexpected a performed action was, given what object the prop was imagined (or pretended) to be.* For example, if the prop had been imagined to be a mop, it might have been used to clean the floor, which would be unsurprising; however, if it was used to row a boat, that would be surprising. *Quality* was defined as *how smooth, and recognizable the action was.* As mentioned earlier, creativity was defined by the user, and they were asked to strictly use that same definition when evaluating the actions for *perceived creativity* later on. They were also asked about memorable reasons why the marked memorable actions more or less creative than others.

Results

The results from the experimental and control conditions for novelty, unexpectedness, and quality are displayed in table 3.1. The table shows the mean, median, and standard deviation for participant accuracy across each of the ten experimental and control pairings in their task. The accuracy is calculated to mean whether the model accurately predicted their response or not (or vice versa). The results for the user-defined creativity evaluations across all three tasks are similarly displayed in table 3.2 and show a similar set of data.

Statistical hypothesis testing was done in order to measure the statistical significance of our findings about the relative perceptual accuracy between the computational models of novelty, unexpectedness, and quality. The null hypotheses ($H_{0,1}$ to $H_{0,7}$) stated that there were no significant differences for results from a specific question. The alternate hypotheses (H_1 to H_7) for each of the questions about novelty, unexpectedness (two hypotheses

Table 3.1: Mean, Median, and Standard Deviation for participant accuracy.

	High vs Low			Med. vs Med.		
	μ	M	σ	μ	M	σ
Novelty	0.4582	0.5	0.1595	**0.5564**	0.5	0.1619
Object Surprise	0.4818	0.5	0.1622	**0.5364**	0.5	0.1682
Action Surprise	0.4818	0.5	0.1645	**0.54**	0.5	0.1355
Quality	**0.4473**	0.5	0.1464	0.4	0.4	0.1427

Table 3.2: Mean, Median, and Standard Deviation for perceived creativity results from novelty (N), surprise (S), and quality (Q) tasks respectively.

	High vs Low			Med. vs Med.		
	μ	M	σ	μ	M	σ
Creativity (N)	**0.4836**	0.5	0.1719	0.4309	0.5	0.1538
Creativity (S)	0.48	0.5	0.1899	**0.5327**	0.5	0.1667
Creativity (Q)	**0.4491**	0.5	0.2045	0.4455	0.5	0.1942

each about the object surprise and action surprise respectively), quality, and user-defined creativity (each task separately asked them about user-defined creativity leading to 3 alternate hypotheses for each question about it). After finding non-normality in the distributions of responses using a Shapiro-Wilk [216] test, the non-parametric, repeated measures, Wilcoxon Signed-Rank Test [218] was used to determine significance between the experimental and control condition. The results from significance testing are in table 3.3 along with the effect size to be interpreted as small effect > 0.1, medium effect > 0.3, and large effect > 0.5.

Table 3.3: Wilcoxon Signed-Rank Test for novelty, surprise, and quality task results. Bold significant at p < 0.5. Shows P-values and effect size (ϕ).

	Novelty		Object Surprise		Action Surprise		Quality	
	p	ϕ	p	ϕ	p	ϕ	p	ϕ
Total	**0.0005**	**0.466**	0.0731	0.242	**0.0458**	**0.270**	0.1203	0.210
Creativity	0.1123	0.214	0.0793	0.237	-	-	0.9024	0.017

These results show that observers could not reliably recognize the predictions of the

computational models for evaluating creativity in terms of novelty, unexpectedness, and quality over actions that were performed by a robot character. The hypothesis for this approach was that there would be significant differences in the recognition accuracy for the different pairs (i.e., between the high-low pair and the medium-medium pair). It was expected that if the predictions matched human perceptions for these properties of actions, then the high-low pairing would be more obviously and reliably comparable than the medium-medium pair. However, from the results, this did not turn out to be the case. It can be seen from table 3.3 that the only significant differences in the distributions of responses were in the recognition accuracies between high-low and medium-medium for Novelty and Action Surprise. However, these effects were both in the opposite direction of significance than we had hoped for. The effect size $\phi_{Novelty}$ indicates medium effect size while the $\phi_{ActionSurprise}$ indicate a small effect. The reported accuracies for all pairs were close to the 0.5 random selection baseline, though they were consistently slightly lower accuracy than that baseline for most measures in the high-low pairing and consistently slightly above that baseline for medium-medium measures.

This is a negative result from the context of validating the agent's models for creativity evaluation in terms of perceptual similarity with human observers. There could be many reasons for this result, including the knowledge and expectation disparity, the difference between the conceptual representation of the action and the actual performance of it, the difference between experiencing the system as a participant and an observer, and even possible errors like incorrect parameter configurations in the system. Some of these potential reasons will be discussed in the following section.

Discussion

This study aimed to validate that the ratings from the creativity evaluation models perceptually matched a human observer's ratings. The study provided evidence that the creativity evaluation models did not succeed at matching the human observers' perceptions of con-

cepts such as novelty, unexpectedness, quality, or creativity. It is possible that this is always going to be the case until the system gets enough knowledge, experience, and builds expectations to match humans. Alternatively, it is possible that in order to improve the effectiveness of these models, the participant's experiences and expectations have to be teased out through personalization and modeling. This is the approach taken by Grace, Maher, Mohseni, and Pérez [219] and points to a potential future direction to take for this work.

Another perspective perhaps, as the saying goes, is to consider that, "all models are wrong, but some models are useful." Therefore, a more fundamental question might be whether the models are evaluating some useful aspect of an improvisational collaborator's or audience member's experience to make a meaningful difference in the agent while improvising, in terms of their perceptions of enjoyment, creativity, and coherence. If the models are doing this already through guiding the agent's action selection but can't match the quality of experience, they are evaluating to overloaded concepts like novelty, surprise, value, or creativity that may be acceptable. Results from the next set of evaluation studies would indicate that this is the case since the creativity evaluation models are being used to deliver identifiable creative arc negotiation and study participants seem to prefer sessions with creative arc negotiation across different criteria. It could also be that the task of comparing individual actions directly in this experiment was too challenging for raters, whereas the longer, session-length task provided more context for them to evaluate similar parameters. More study is required to disentangle what exactly the system's creativity evaluation models measure in this case.

3.5.3 Evaluating Creative Arc Identification with Observers

This experiment is the first of three studies that evaluate the claims made in the thesis statement directly (RQ5). The aim of the experiment is to understand whether observers of an improvised performance can correctly identify trends in various parameters according to the creative arc used to drive action selection in each experimental condition. The results of

this study, if successful, would serve to provide evidence (in concert with other results) that a) CARNIVAL agents successfully perform creative arc negotiation and that people can recognize that as well as that b) the improvised performances resulting from using creative arc negotiation within CARNIVAL are meaningfully different enough to allow people to recognize those differences in the performances (in terms of the different arcs they perceive) and thus, at least partially address the improvisational action selection problem.

Methodology

A survey-driven, non-expert, observer-rating study was performed in an attempt to evaluate whether we had successfully created an embodied agent architecture that enables an agent to negotiate a given creative arc while performing movement improv with non-experts. This was performed in combination with a pilot, in-person, non-expert, participant/interactor-rating, laboratory study described later (see section 3.5.5). Since this was an observer-rating study, it was designed to measure the degree to which observers could correctly identify the nature of the creative arc in different improvised performance sessions, where the agent's action selection was performed by creative arc negotiation, through observation.

The three creative arcs used in the sessions for comparison were respectively *rising*, *falling*, and *level* arcs. The values for each creative arc (with each creative space dimension in the closed interval $[0.0, 1.0]$) were as follows. The rising arc had a linearly rising arc over five turns of the props game, each ranging from $< 0.0, 0.0, 0.5 >$ to $< 1.0, 1.0, 1.0 >$. The falling arc had the opposite arc from $< 1.0, 1.0, 1.0 >$ to $< 0.0, 0.0, 0.5 >$. Finally, the level arc always had the scores $< 0.5, 0.5, 0.5 >$. Future work could also compare level arcs with values at $< 1.0, 1.0, 1.0 >$ or $< 0.0, 0.0, 0.5 >$ to see how those constantly maximum and minimum values in the creative space to the current set of arcs.

One hundred non-expert raters on Amazon Mechanical Turk were asked to watch videos of three different sessions between a researcher and the agent (which was controlled by creative arc negotiation). For each video, they were then asked to choose whether a given

property of the performance was rising, falling, or level (i.e., with one-third or 33.33% probability of selecting correctly at random). Each video was taken of the agent controlled by one of the three creative arcs described above. The different qualities that were asked of them were *novelty*, *object surprise*, *action surprise*, *quality*, and *user-defined creativity*. All raters were given the definition of each property in the question (as defined in section 3.5.2 previously) except user-defined creativity, which they were made to define before the rating task started. The study hypotheses were that there would be significant differences between the different arcs in terms of relative recognition rates among all participants.

Results

The relative percentages of participants who correctly identified the option for each rated property of the video session that observers were asked to identify as rising, falling, or level are presented in table 3.4. It is important to note that in this task, a random baseline would score 33.33% of its choices correctly since there are three choices from which to choose an answer.

Table 3.4: Relative recognition percentages between arc types in creative arc identification task. Bold is higher between pairs.

	Rising		Falling		Level	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Total	$\mathbf{56.41\%}$	43.59%	44.95%	$\mathbf{55.05\%}$	38.29%	$\mathbf{61.71\%}$
Novelty	$\mathbf{57.14\%}$	42.86%	37.14%	$\mathbf{62.86\%}$	20.95%	$\mathbf{79.05\%}$
Object Surprise	$\mathbf{53.33\%}$	46.67%	44.76%	$\mathbf{55.24\%}$	34.29%	$\mathbf{65.71\%}$
Action Surprise	47.12%	$\mathbf{52.88\%}$	37.14%	$\mathbf{62.86\%}$	42.86%	$\mathbf{57.14\%}$
Quality	$\mathbf{73.08\%}$	26.92%	$\mathbf{61.90\%}$	38.10%	$\mathbf{60.00\%}$	40.00%
Creativity	$\mathbf{51.43\%}$	48.57%	43.81%	$\mathbf{56.19\%}$	33.33%	$\mathbf{66.67\%}$

A Chi-Squared Test of Independence was used to calculate whether there were significant differences in relative recognition rates between the different arcs among all participants. The null hypotheses ($H_{0,1}$ to $H_{0,5}$) for the five questions were that there was no significant difference between the distributions of responses for each arc. The alternate

hypotheses (H_1 to H_5) stated that significant differences did exist between the distributions of responses for the three creative arc-driven performances. The results can be seen in table 3.5. A further Chi-Square Goodness of Fit test was performed to evaluate whether there were significant results between correctly vs. incorrectly identifying the direction of the arc for the given property. The null hypotheses ($H_{0,6}$ to $H_{0,10}$) for the five questions were that there was no significant difference between the distributions of responses identifying the arc for each property. The alternate hypotheses (H_6 to H_10) stated that significant differences did exist between the distributions of responses identifying the arc for each property. The results for each rated property of the session from the Chi-Square Goodness of Fit test are presented in table 3.6.

Table 3.5: Chi-Square test of independence for creative arc identification task. Bold significant at p < 0.5. ϕ is effect size.

	X^2	p	ϕ
Total	$\mathbf{35.3675}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.150}$
Novelty	$\mathbf{29.1731}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.304}$
Object Surprise	$\mathbf{7.7514}$	$\mathbf{0.0207}$	$\mathbf{0.157}$
Action Surprise	2.1444	0.3423	0.083
Quality	4.5762	0.1015	0.121
Creativity	$\mathbf{7.0778}$	$\mathbf{0.029}$	$\mathbf{0.150}$

Table 3.6: Chi-Square goodness of fit for creative arc identification task arcs. Bold significant at p < 0.5. ϕ is effect size. Object Surprise and Action Surprise contracted for space.

	Rising			Falling			Level		
	X^2	p	ϕ	X^2	p	ϕ	X^2	p	ϕ
Total	$\mathbf{125.28}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.490}$	$\mathbf{31.89}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.247}$	$\mathbf{5.79}$	$\mathbf{0.01608}$	$\mathbf{0.106}$
Novelty	$\mathbf{26.79}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.505}$	0.69	0.40763	0.081	$\mathbf{7.24}$	$\mathbf{0.0071}$	$\mathbf{0.263}$
O Surprise	$\mathbf{18.90}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.424}$	$\mathbf{6.17}$	$\mathbf{0.013}$	$\mathbf{0.242}$	0.04	0.836	0.020
A Surprise	$\mathbf{8.89}$	$\mathbf{0.0029}$	$\mathbf{0.292}$	0.69	0.4076	0.081	$\mathbf{4.29}$	$\mathbf{0.0384}$	$\mathbf{0.202}$
Quality	$\mathbf{73.92}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.843}$	$\mathbf{38.57}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.606}$	$\mathbf{33.60}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.566}$
Creativity	$\mathbf{15.47}$	$\mathbf{0.00008}$	$\mathbf{0.384}$	$\mathbf{5.19}$	$\mathbf{0.0228}$	$\mathbf{0.222}$	0	1	0

The differences in relative percentages of participants correctly identifying the rising,

falling, and level creative arcs and the statistical hypothesis testing show that for the specific sets of significantly differing properties of the performances (i.e. for total response, novelty, object surprise, and creativity), the videos of sessions with rising and falling arcs could be identified reliably (note that 66% is twice the expectation for a random baseline guess in this task due to the three options present for every question). It also shows the notable trend across all significantly different properties, that recognition accuracy for level arcs is consistently and significantly as bad as random guessing. However, the effect sizes to determine statistical differences across the different types of arcs are predominantly in the small effect range. The Goodness of Fit results, however, show medium and large effects consistently. They indicate that the properties of the video session that we were interested in tracking over the course of the performance were reliably different from the random guessing baseline recognition accuracies. This meant in general that these properties relating to the definition of creativity used in this system were reliably identifiable with rising arcs, less so with falling arcs, and difficult to identify for level arcs. These results are also discussed in more detail in section 3.6.1, especially their relationship to the claims made in my thesis statement.

3.5.4 Evaluating Creative Arc Preferences with Observers

This observer evaluation study is the second of three that directly evaluates claims made in my thesis statement (RQ5). This evaluation experiment aimed to understand the effect of creative arc negotiation on the perceived enjoyment, agent creativity, and coherence on observers as compared to a random action selection alternative. The results of this study, if successful at increasing these perceptions for observers, would serve as evidence (along with other results) that a) using creative arc negotiation successfully addresses the improvisational action selection problem, b) at least for observers, using creative arc negotiation increases perceptions of enjoyment, agent creativity, and coherence.

Methodology

A survey-driven, non-expert, observer-rating study was performed in an attempt to evaluate whether an embodied agent that could negotiate a creative arc while performing movement improv with non-experts was successfully able to increase audience and user perceptions of enjoyment and agent creativity. Since we were using an observer-rating study, we designed it to measure the degree to which observers would prefer videos of improvised sessions between a researcher and an agent that was controlled by either creative arc negotiation or random sampling from the agent's sample space (though still using affordance-based action variant generation, just not the other two main components).

The two conditions compared in the study were a *creative arc negotiating agent* and a *random sampling agent*. Additionally, the creative arc negotiating agent in the three respective videos was controlled by three different creative arcs — *rising*, *falling*, and *level* creative arcs. These followed the same values for the arcs in the creative space as in the Creative Arc Identification study (see section 3.5.3).

One hundred non-expert raters on Amazon Mechanical Turk were asked to watch videos of two different sessions between a researcher and the agent, who was either controlled by creative arc negotiation or performing random sampling. For each video, they were then asked to choose whether they preferred the one on the left or the one on the right in a forced-choice configuration based on the given property of the performance. Each video had a baseline random probability of being selected half the time (or 50%). The different qualities that they were asked to compare were enjoyment, user-defined creativity, and coherence. Before the study, participants were made to define creativity before the rating task started and asked to restrict themselves to that definition of creativity during the task. The study hypotheses were that there would be significant differences between the choice of creative arc vs. no arc action selection as well as between the different arcs in terms of which ones were preferred as compared to the no arc condition.

The initial study was repeated with the same methodology using videos with just the

agent's turns spliced together from the original video (the researcher's actions were removed). This was done to mitigate any bias in the results from the human's actions, i.e., in case the human's actions contributed positively or negatively to the degree of preference for a certain type of response from participants. This time, however, the sample size was increased to be one hundred twenty participants.

Results

The percentage of participants who chose the creative arc-negotiating performances vs. the random sampling performances for each rated property of the video session are presented in table 3.7. The properties that participants were asked about included which session they enjoyed more, in which session did the agent seem more creative, and which session was more coherent. It is important to note that in this task, a random choice baseline would score 50% since there are only two choices from which to choose an answer. Clear trends are present from the table in the relative percentages of preference for creative arc negotiating performances and random sampling performances.

Table 3.7: Relative preferences between an arc condition and a no arc condition in creative arc comparison task. Bold is higher between pairs.

	Rising		Falling		Level	
	Arc	No Arc	Arc	No Arc	Arc	No Arc
Total	**69.35%**	30.65%	**65.06%**	34.94%	28.12%	**71.88%**
Enjoyment	**63.46%**	36.54%	**62.50%**	37.50%	28.57%	**71.43%**
Agent Creativity	**64.08%**	35.92%	**56.73%**	43.27%	31.73%	**68.27%**
Coherence	**80.58%**	19.42%	**75.96%**	24.04%	24.04%	**75.96%**

A Chi-Squared Test of Independence was used to calculate whether there were significant differences in the proportion of creative arcs selected between the different arcs among the participants. The null hypotheses ($H_{0,1}$ to $H_{0,5}$) for the five questions were that there was no significant difference between the distributions of responses for each arc. The alternate hypotheses (H_1 to H_5) stated that significant differences did exist between the

distributions of responses for the three creative arc-driven performances. The results can be seen in table 3.8. For properties that found significance during the preceding Chi-Square Test of Independence, a further Chi-Square Goodness of Fit test was performed to evaluate whether there were significant results in the preference of one video performance vs. the other with respect to specific properties about which the question was asked (e.g., about which one was more enjoyable). The null hypotheses ($H_{0,6}$ to $H_{0,10}$) for the five questions were that there was no significant difference between the distributions of responses for each arc to the expected outcome in each case. The alternate hypotheses (H_6 to H_10) stated that significant differences did exist between the distributions of responses for the three creative arc-driven performances with respect to the expected outcomes. The results for each rated property of the session from the Chi-Square Goodness of Fit test are presented in table 3.9.

Table 3.8: Chi-Square test of independence for creative arc comparison task. Bold significant at p < 0.5. ϕ is effect size.

	X^2	p	ϕ
Total	$\mathbf{129.27}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.372}$
Enjoyment	$\mathbf{33.09}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.325}$
Agent Creativity	$\mathbf{23.86}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.277}$
Coherence	$\mathbf{85.35}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.524}$

Table 3.9: Chi-Square goodness of fit for creative arc comparison task arcs. Bold significant at p < 0.5. ϕ is effect size. Agent Creativity contracted for space.

	Rising			Falling			Level		
	X^2	p	ϕ	X^2	p	ϕ	X^2	p	ϕ
Total	$\mathbf{46.45}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.387}$	$\mathbf{28.32}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.301}$	$\mathbf{59.97}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.438}$
Enjoyment	$\mathbf{7.54}$	$\mathbf{0.00604}$	$\mathbf{0.269}$	$\mathbf{6.5}$	$\mathbf{0.01079}$	$\mathbf{0.250}$	$\mathbf{19.29}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.429}$
Creativity	$\mathbf{8.17}$	$\mathbf{0.00427}$	$\mathbf{0.282}$	$\mathbf{1.89}$	$\mathbf{0.013}$	$\mathbf{0.135}$	$\mathbf{13.89}$	$\mathbf{0.00019}$	$\mathbf{0.365}$
Coherence	$\mathbf{38.53}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.612}$	$\mathbf{28.04}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.519}$	$\mathbf{28.04}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.519}$

The results of this experiment suggested there were significant, reliably detectable preferences for the creative arc negotiation-driven agents, at least for observers viewing videos

of performances and comparing it with a random sampling agent baseline. All three properties (enjoyment, agent creativity, and coherence) were significantly different and showed effect sizes ranging from small to large along with positive (and desirable) differences. The effects for rising and falling arcs (with the effect stronger in general for rising arcs) showed that coherence was the most improved with agent creativity and enjoyment following closely behind.

The results from the repeat performance of this study with just the agent's actions spliced together in the video were analyzed exactly the same way as the previous case. The results for recognition accuracies across arcs can be seen in table 3.10. After performing statistical significance testing, the results can be seen in tables 3.11 and 3.12.

Table 3.10: Relative preferences between an arc condition and a no arc condition in creative arc comparison task with only agent turns (no human turns). Bold is higher between pairs.

	Rising		Falling		Level	
	Arc	No Arc	Arc	No Arc	Arc	No Arc
Total	**84.44%**	15.56%	**69.08%**	30.92%	21.85%	**78.15%**
Enjoyment	**86.67%**	13.33%	**70.83%**	29.17%	23.53%	**76.47%**
Agent Creativity	**73.33%**	26.67%	**63.03%**	36.97%	25.21%	**74.79%**
Coherence	**93.33%**	6.67%	**73.33%**	26.67%	16.81%	**83.19%**

Table 3.11: Chi-Square test of independence for creative arc comparison task with only agent turns (no human turns). Bold significant at $p < 0.5$. ϕ is effect size.

	X^2	p	ϕ
Total	$\mathbf{314.01}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.54}$
Enjoyment	$\mathbf{107.75}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.548}$
Agent Creativity	$\mathbf{61.65}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.415}$
Coherence	$\mathbf{158.51}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.664}$

The results for the repeated observer study with just footage of the agent taking its turns in order showed an even stronger effect in the same direction as the previous study. This allowed us to remove the effect of the human on the observed effects. It also allowed us

Table 3.12: Chi-Square goodness of fit for creative arc comparison task arcs with only agent turns (no human turns). Bold significant at $p < 0.5$. ϕ is effect size. Agent Creativity contracted for space.

	Rising			Falling			Level		
	X^2	p	ϕ	X^2	p	ϕ	X^2	p	ϕ
Total	$\mathbf{170.84}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.689}$	$\mathbf{52.28}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.382}$	$\mathbf{113.17}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.563}$
Enjoyment	$\mathbf{64.53}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.733}$	$\mathbf{20.83}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.417}$	$\mathbf{33.35}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.529}$
Creativity	$\mathbf{26.13}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.467}$	$\mathbf{8.08}$	$\mathbf{0.00449}$	$\mathbf{0.261}$	$\mathbf{29.25}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.496}$
Coherence	$\mathbf{90.13}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.867}$	$\mathbf{26.13}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.467}$	$\mathbf{52.45}$	$\mathbf{< 10^{-5}}$	$\mathbf{0.664}$

to address suspicions of researcher bias, from the previous iteration of the study, in terms of implicitly shaping the videos for evaluation. This is a valid concern since it is a co-creative performance with creative responsibilities falling on the shoulders of both human and computer improviser. It would be natural for there to be researcher bias or error in constructing the comparison videos. However, the results from the repeated iteration of the study lay those concerns to rest and improve on the previous results in terms of effect size and increased preference for the creative arc negotiation versions of the system. Additional discussion is also found on this topic in section 3.6.1 and how it relates to the claims in my thesis statement.

3.5.5 Evaluating Creative Arc Improvisation with In-Person Participant Pilot

This pilot, participant study is the final of three that directly evaluates claims made in my thesis statement (RQ5). The aim of this study was to repeat the previous two studies conducted for observer ratings and understand how the use of creative arc negotiation would affect participants in terms of their perceptions of enjoyment, agent creativity, and coherence. If successful at increasing these perceptions for participants, the results would provide further evidence for a) the ability of creative arc negotiation to address the improvisational action selection problem and b) that using creative arc negotiation increases perceptions of enjoyment, agent creativity, and coherence.

Methodology

A pilot, non-expert, participant-rating, laboratory study was conducted to get quantitative and qualitative feedback about the experience of interacting with the CARNIVAL architecture in the Robot Improv Circus installation. The study was aimed at understanding whether participants/interactors could 1) identify whether the qualitative trends of different properties of an improvised performance matched the specific creative arc that the agent used to guide action selection and 2) whether they preferred the subjective experience of interacting with the agent when it was using creative arc action negotiation or random sampling from its action space to guide action selection. Additionally, since the study was intended as an initial small-scale pilot study, importance was given to both the quantitative responses that were received and the semi-structured interview content.

A total of 18 participants were recruited for the initial pilot study in two batches of 6 and 12 from a non-expert student population. However, due to differences in the tasks performed by the two batches of participants, the number of responses for the tasks and individual questions in the study differed between either 12 or 18 (these differences will be noted when reporting results). Participants were first given a pre-study experience questionnaire to complete. They were then given an opportunity to get familiar with how to use the VR system and the specific installation through a tutorial VR environment and a set of trial rounds for the installation, respectively. Participants were next placed into one of three groups at random and continued on to complete the two study tasks. The groups in which each set of participants were placed will be explained in the individual contexts of the two study tasks later. Finally, the study concluded after participants were debriefed and compensated for their participation.

The first of the experimental tasks was creative arc comparison. For this task, the participant was asked to perform two rounds of improvisation with the agent. The agents were in different action selection conditions according to the specific task group that the participant was assigned to for each session. After improvising with the agent twice, the

participant was asked to compare the two sessions through a survey and a semi-structured interview. The groups that participants were assigned to for this task were 1) rising arc vs. no arc/random sampling, 2) falling arc vs. no arc/random sampling, and 3) level arc vs. no arc/random sampling. The ordering for conditions within each group was randomized as well.

The session comparison questionnaire for these tasks asked the following two to three questions depending on which batch of the pilot was being run. "1) Which of the sessions did you enjoy more?" "2) In which of the sessions would you say your partner was more creative overall?" "3) Which of the two sessions would you say was more logical overall?" The first two questions received 18 responses, while the third question received 12 responses. Additionally, for the second question, participants were asked to reflect on their own definition of creativity before completing this questionnaire and were asked to clarify that definition during the semi-structured interviews. For both questions, participants could select between the response options — session 1, session 2, both equally, and not sure. During the semi-structured interview, participants were asked questions to clarify their definition of creativity used in the previous questionnaire, memorable reasons or examples of interactions that led to them picking one session over the other for creativity or enjoyment, and other reasons why they preferred one session over the other. Participants were also asked for open-ended feedback on the interaction, experience, or other aspects of the sessions.

The second experimental task was creative arc identification. For this task, the participant performed two rounds of improvisation with the agent and answered questions after each session about their experience. Each session was evaluated with a questionnaire and a semi-structured interview. In this task, the three groups that participants were assigned to at random were 1) rising arc vs. level arc, 2) falling arc vs. level arc, and 3) rising arc vs. falling arc.

The session evaluation questionnaire for the arc identification task included the fol-

lowing kinds of questions. 1) Whether novelty, object surprise, action surprise, and user-defined creativity (with definitions for each property except user-defined creativity presented to them as defined in section 3.5.2) increased, decreased, or stayed the same over time. 2) The level of overall quality (with the definition for quality presented to them as defined in section 3.5.2) for the agent's actions performed from very low quality to very high quality on a five-point Likert scale. 3) Their level of agreement, on a five-point Likert scale from 'strongly disagree' to 'strongly agree,' to the statement, "Over time, my enjoyment of the experience increased." The semi-structured interview questions involved asking them what their definition of creativity was that they used for the questionnaire, whether there were memorable reasons or examples to explain their responses to the various performance trend questions. Additionally, participants were also asked for open-ended feedback on the interaction, experience, or other aspects of the sessions.

Results

The results from the different questionnaires for the two study tasks are summarized and presented in tables 3.13 and 3.14 as well as table 3.15. The first table summarizes the relative differences in four possible preferences (creative arc, no arc, both, or neither) between the two conditions compared in the task (creative arc and no arc). The second table shows the result of performing a Chi-Square goodness of fit test on the combined data for comparing creative arc sessions against no arc sessions. Note that for both analyses of the creative arc comparison task, the sample size was too small to split the conditions feasibly according to arc type. The sample size for the creative arc identification study was not large enough to show statistically significant differences in the distributions of results across arcs. The initial results from the study will be expanded in the future, and the evidence will be re-reviewed after increasing the sample size to get more confident results. Discussion of the implications of this study as it relates to the thesis statement for this dissertation can be found in section **??**.

Table 3.13: Relative preferences for arc, no arc, both, or neither between a creative arc session and a no creative arc (random action selection) session in the participant-rating creative arc comparison task. Bold is highest for the row. N is sample size.

	Creative Arc	No Arc	Both	Neither	N
Enjoyment	**38.89%**	33.33%	27.78%	0%	18
Agent Creativity	**55.56%**	27.78%	5.56%	11.11%	18
Coherence	**58.33%**	8.33%	25%	8.33%	12

Table 3.14: Chi-Square goodness of fit between combined arc and no arc sessions for the participant rating creative arc comparison task. Bold significant at $p < 0.5$. ϕ is effect size. N is sample size.

	X^2	p	ϕ	N
Enjoyment	6.44	0.09188	0.598	18
Agent Creativity	**10.89**	**0.01234**	**0.778**	**18**
Coherence	**8**	**0.04601**	**0.816**	**12**

The qualitative results from the semi-structured interviews are still being analyzed. However, they have already resulted in some useful questions for guiding the future directions of this research. This includes questions about the implementation of creativity evaluation models in the future and questions about the installation and interaction design for an improvisational agent-based interactive installation. Some of the initial questions that have arisen from reflection about them so far are as follows. 1) How could the interaction design for the user change to better enable access to framing information that non-experts can understand to explain how the agent's creative process currently works? This question became relevant when discussing the participants' mental models of how the agent arrived at a particular response. Some initial ideas on this topic can be seen in [220] 2) While some existing works have shown that framing does not have to be truthful, what are the performative affordances for visualizing the system's actual reasoning process to an audience versus a post-hoc explanation? This question arose as a follow up to the previous question during discussion within the research team. 3) Given how differently people seem to experience the outputs of the creativity evaluation models, does the model need to local-

Table 3.15: Session creative arc identification results for participants in pilot study. Four participants P1 - P4 per creative arc type.

Arc Type + Question	P1	P2	P3	P4
Rising Q1: Novelty	Rising	Level	Rising	Level
Falling Q1: Novelty	Rising	Level	Rising	Rising
Level Q1: Novelty	Rising	Rising	Level	Falling
Rising Q2: Object Surprise	Rising	Level	Level	Falling
Falling Q2: Object Surprise	Falling	Level	Level	Rising
Level Q2: Object Surprise	Rising	Rising	Level	Rising
Rising Q3: Action Surprise	Rising	Rising	Rising	Falling
Falling Q3: Action Surprise	Falling	Rising	Level	Falling
Level Q3: Action Surprise	Falling	Rising	Rising	Rising
Rising Q4: Quality	Moderate	Low	Moderate	Very Low
Falling Q4: Quality	Low	Moderate	Low	Very Low
Level Q4: Quality	Low	Moderate	Low	Low
Rising Q5: Creativity	Rising	Level	Level	Rising
Falling Q5: Creativity	Level	Level	Rising	Rising
Level Q5: Creativity	Rising	Rising	Level	Rising
Rising Q6: Enjoyment Increased	Agree	Agree	Neither	Strongly Agree
Falling Q6: Enjoyment Increased	Agree	Neither	Agree	Strongly Agree
Level Q6: Enjoyment Increased	Agree	Strongly Agree	Agree	Agree

ize its computation for the current user to be effective? Initial work in this area from [219] suggests that it could be an effective strategy when dealing with the variety of human experiences. However, this is difficult to implement when the agent's expertise is much lower than the human collaborator's experience level (see section 3.3.6). 4) Users had conflicting feedback on why something was good or bad. At least some of the conflict seems to arise from who is being modeled as the experiencing agent for the creativity evaluation models. Therefore, the following question arose as a potential direction for future exploration. How can the agent model and combine multiple perspectives (e.g., audience vs. interactor) for computationally evaluating creativity? 5) Finally, the discussions with participants resulted in a fundamental question about the creative goals for the installation. Ultimately, who is this installation for the audience or the interactor, and how should the design change to make that clearer? Perhaps if framed for the participants as a performative, rather than,

participatory, installation, the expectations for the experience would be clearer to a non-expert (in contrast to a professional improviser who would subtly prioritize the audience's entertainment above other considerations). These questions aren't necessarily easy to answer, but they point to future directions for exploration of this work in the installation and experience design space.

3.6 Discussion

The preceding set of evaluation studies with observers and participants of improvised performance from the installation aimed to evaluate the claims in my thesis statement. In this discussion section, results from all three evaluation experiments are synthesized together in the following subsection. This is followed by a summative discussion subsection that compares the evidence from these studies against the claims in my thesis statement to draw conclusions its validity.

3.6.1 Evaluating Improvisation Using Creative Arc Negotiating with Observers and Participants

The aims for the three creative arc improvisation evaluation studies was to understand better if the creative arc negotiation that the agent was performing actually made a difference to users of the improvised performance whether as observers or participants. For example, could observers and installation participants actually understand what kind of creative arc an agent used in a given performance? Additionally, aside from identifying a difference between these arcs, would observers or installation participants actually prefer sessions of improvisation guided by this form of action selection?

The strong results from both of the studies involving observers comparing videos of improvised performances with different creative arcs or performances with and without creative arc negotiation were particularly notable to me for several reasons. Firstly, the results from both studies suggested that despite the perceptual (or conceptual) mismatch between observer ratings and ratings from the agent's creativity evaluation models about

actions within the domain (as evidenced by results in section 3.5.2), the agent was able to produce a meaningful impact to an observer's experience based on the presented experimental variations. Secondly, the result that observers strongly preferred sessions where the improv was guided by creative arc negotiation for action selection in comparison to random sampling of the agent's action space in terms of enjoyment, agent creativity, and coherence was a fundamental demonstration that creative arc negotiation did, in fact, address the improvisational action selection problem, at least for observer ratings of the improvised performances. Thirdly, the result that observers could identify differences between all three kinds of arcs and preferred them (in comparison to random sampling) to different degrees indicated that certain arcs might be better suited for improvised performance than others, at least for observers. The existence of these preferences makes sense given the large body of literature on the presence of specific (and limited) sets of arcs across narratives for dramatic tension and plot [221, 222, 223] or character affect [224]. Future work might have creative arcs over longer sessions that rose and then fell accordingly. Fourthly, the result that observers had a significant preference for random sampling over a flat arc was initially unexpected. However, this could provide additional support for my dissertation's thesis that a continuously evolving experience over time results in increased enjoyment, coherence, and perceptions of creativity, at least for observers and for the particular medium-level arc that was evaluated.

The results from the in-person installation participant ratings indicated similar results for creative arc preference for agent creativity and coherence but not for enjoyment (which needed more data and study to show a significant preference conclusively in either direction). Creative arc identification did not yield significant results either, possibly due to the much smaller sample size and the resulting loss of statistical significance. The sample size was also too small to split the data according to arc type for the creative arc preference/comparison studies, and so the results are presented purely in comparison of creative arc negotiation (with any arc type) and no arc (i.e., random sampling) action selection.

The results from the two observer studies demonstrate strong effects in the hypothesized direction and provide strong evidence to verify the claims made in my thesis statement. The claims were also partially supported (to different degrees) by the results of the pilot in-person installation participant studies (conclusively supporting it, despite the small sample size, for increases in participant perceptions of agent creativity and coherence but requiring more study/data to do the same for participant enjoyment). This is quite likely due to the small sample size; however, there is some evidence from [225, 226] that suggests that interactors in the midst of an ephemeral improvisational experience may have trouble keeping track of longer-term effects. Kelso, Weyhrauch, and Bates [225] describes this as a positive effect of interactive narrative experiences, saying that interactive narratives do not need to preserve narrative coherence as strongly as other forms of narrative since participants will not be able to keep track of these longer-term causal links in any case. However, perhaps the converse also applies to methods that attempt to show meaningful differences in user experience for participants between different experimental interventions that employ system ablation. Kelso, Weyhrauch, and Bates's finding also implies that it could be more difficult to demonstrate working interventions in the user's experience based on longer-term effects in ephemeral, but temporally extended, participatory experiences like movement improv.

3.6.2 Evaluating Thesis Statement

My thesis statement states that "embodied agents that address the improvisational action selection problem using 'creative arc negotiation' increase perceptions of enjoyment, agent creativity, and coherence in both observers and participants while performing movement improv with non-experts." The preceding section of this chapter presented a set of experiments for evaluating the claims in that thesis statement. Throughout this section, I discuss the results of those experiments and draw conclusions about the validity of my thesis statement.

The agent architecture CARNIVAL was designed directly to enable embodied agents to use creative arc negotiation (albeit tailored to the Props game domain in some aspects). Additionally, the results of the observer-rating creative arc identification study showed that, at least for observers, the improvisational partner was able to select actions to follow a creative arc over the course of the improvised performance. This result was not fully replicated with statistical significance in the laboratory participant-rating evaluation due to the sample size constraints on the study. However, at least for observers, my results validate the claim that **embodied agents within CARNIVAL successfully demonstrate the usage of creative arc negotiation.**

It is valid to claim that the agent architecture enabled improvisational agents to select actions in near real-time, at least within the constraints of the chosen domain to facilitate successful improvisation, since both participants were able to perform alongside the agent and observers rated the improvisational performances higher than 'no arc' action selection. Additionally, different versions of the creative arc negotiation agents (with different creative arcs) showed accurately identifiable differences in observer experiences and different versions of the agents (creative arc negotiation vs. no arc action selection) showed consistently different preferences for those experiences (at least for observers). Finally, the improvised sessions with creative arc negotiation were rated better for coherence than an alternative that did not use creative arc negotiation. Therefore, it is valid to claim that the **improvisational action selection problem is successfully addressed by creative arc negotiation as an approach, with strongly positive evidence from observers but with initially positive evidence from participants that needs further study to become more conclusive.**

It has also been shown through public demonstrations/exhibitions, feedback from participants, and the participant-rating in-person study of creative arcs that the Robot Improv Circus is an installation that successfully allows participants to perform movement improv with an embodied improvisational agent. Further audience members can successfully

watch, cheer, and give supportive feedback to participants within the installation through the design of the installation. Observers also prefer the version of the installation that performs creative arc negotiation in terms of enjoyment, agent creativity, and coherence. These results are partially replicated for in-person participants as well (with more study needed to show a conclusive preference for enjoyment ratings). Therefore, **embodied agents addressing the improvisational action selection problem using creative arc negotiation (as shown in the preceding paragraph) can successfully perform movement improv with non-experts. Furthermore, this interaction with the installation experience can occur as either a participant or an audience member.**

The results from the creative arc identification and creative arc preference/comparison studies show that, at least for observers, there is conclusive evidence that action selection using creative arc negotiation is preferred over an alternative that used random sampling action selection in terms of observer perceptions of enjoyment, agent creativity, and coherence. This evidence was even stronger when the human partner's actions were removed from the task, and only the agent's actions were evaluated. Similarly, for participants of the installation, initial evidence showed that perceptions of agent creativity and coherence were higher for improvised performances with creative arc negotiating agents (perceptions of enjoyment require more study to show a preference conclusively in either direction). Therefore, it is **valid to conclusively state that embodied agents addressing the improvisational action selection problem using creative arc negotiation (as shown earlier) can perform movement improv with non-experts so that perceptions of agent creativity and coherence increase for both participants and audience members, but that perceptions of enjoyment only conclusively increase for observers. More study and data is required to show a conclusive increase in perceptions of enjoyment for participants of the installation.**

3.7 Future Work

This chapter of my dissertation presented my research into improvisational agents for movement improv domains using creative arc negotiation and how that affected different aspects of observer and participant experience. This was done with the eventual goal of enabling unconstrained human-computer embodied narrative improvisation in the future. The research presented in this work involved creating a VR interactive installation for human-computer movement improv in the Props game domain, creating an improvisational agent architecture for addressing the improvisational action selection problem in movement improv through creative arc negotiation, and evaluating the installation and architecture according to the claims in my thesis statement. This section is a discussion of the opportunities for future research in this area, the limitations of current research, and future expansions or additions planned for this work.

3.7.1 The CARNIVAL Architecture

The CARNIVAL architecture for controlling improvisational virtual agents while performing movement improv with people was primarily designed to address the improvisational action selection problem by creative arcs and creative arc negotiation to follow a trajectory in its creative space over the course of an improvised performance. This idea was partly inspired by the different types of arcs that are present in various creative domains and provide structure for interpreting and (possibly) generating artifacts from those domains. For example, Aristotelian dramatic arcs [221] or Freytag's Triangle [222] represent trajectories of drama and tension within a narrative. It was also partly inspired by interactive narrative research such as Mateas and Stern's Façade, which successfully used Aristotelian dramatic arcs to control the sequencing of story fragments called 'beats' together and guide user experiences to follow a narrative arc over the course of the resulting interactive narrative.

The CARNIVAL architecture primarily used repeated cycles of learning from demon-

stration and improvisation with that learned knowledge as a way to train a model to perform affordance-based action variant generation the Props game domain. This process enabled the use of a deep generative model for action variant generation from the agent's learned action space. The training requirements for the deep learning model enforced a cyclical batch learning approach to training the model. This was an effective strategy for learning to generate variants from a large, continuous, searchable action space using a batch of demonstrations. However, it did mean that the system could not perform interactive learning over the lifetime of the agent, expanding its experience over time.

Future alternatives to the current approach of repeated learning, performance, and retraining, could instead focus on retraining a copy of the model every N turns and then swapping it with the original model in order to update the agent's action space with new actions perceived from the agent's human collaborator with an online (rather than offline) approach. Instead of retraining and then replacing the old model, another approach could be to use a technique for combining generative networks like Guzdial and Riedl's Combinets [227] to create a blended network for the agent to use. In all such cases, new questions arise that would need to be addressed. Firstly, since the system also learns episodic patterns of action over time and the newly replaced or blended models would almost certainly not map the same conceptual action classes to the same location in the new model's latent space, is there a way to procedurally maintain a mapping between the old and new coordinates of the two models that is updated along with each model update? This would also have to be done without interrupting the flow of execution for the agent.

Secondly, batch learning is done at present in order to allow the agent to replace its current model of the action space with a new model of the action space without interrupting the flow of the improvised performance (since it is done between performances). Since new actions would also need to be segmented correctly and include all the ground truth semantic interpretation annotations for the newly added gestures, there is an open question about how best to obtain this new knowledge without interrupting the flow of the improvised

performance and break the user's immersion. This is another question that would need to be addressed before the CARNIVAL architecture can use an online approach to interactive learning. Acquisition of the segmented gestures and semantically interpreted knowledge is currently made by human annotators segmenting and annotating the newly collected data. Any online learning approach would also need to perform this segmentation and annotation automatically. With enough time, this might be possible using semi-supervised learning. However, given the open-ended nature of the domain and the relative lack of knowledge/experience for the agent in comparison to their human collaborator, it might be best to query them for the appropriate interpretative labels interactively. Users could also be made to segment gestures better through improved interaction design. This particular approach is straightforward future work at the moment. The main source of uncertainty would be the nature of interaction design to best enable this sort of interactive learning approach without pulling participants out of their improvisational experience too much. Speech-based inputs have been considered but haven't been added yet. Speech-based inputs can suffer from recognition accuracy, however, especially for accented speech. In order to better segment the action, participants could be trained to hit a buzzer at the start and end of their turn, rather than only at the end as it currently works. Further segmentation optimization could occur through advances in automated gesture segmentation (like [2]) tailored to the domain.

An alternative approach that could enable the agent to learn from interactors over its lifetime in this work without replacing the existing learning pipeline with other interactive learning approaches (as was the focus of prior work [2]) is to somehow collect more training data in the form of actions (with both gestural and semantic components). Work on this approach has already started and is currently being explored further and integrated. A computer vision pipeline based on [228] has been trained to extract 3D human gestural data from videos. The current plan is to use this on YouTube videos as a way to extract gestural data from them. The system does not work with moving cameras at present, so

video selection will also need to be performed before training or camera movement will need to be detected and compensated for using additional techniques like [229]. Another required improvement is that this pipeline does not currently extract semantic data from the videos in addition to the gestures. Therefore, future work would also investigate the extraction of descriptive natural language tags for gestures in those videos (similarly to [230]). Results could potentially be improved by training the action extraction pipeline on a combination of inputs consisting of matched sets of movies, scripts, and subtitles.

The affordance-based action generation used in this work to conditionally generate action variants based on the physical attributes of the object was designed to encode a learned mapping between the physical attributes of objects and the actions that could be generated with them. This proved to be a successful strategy for partitioning the action selection to appropriate objects as well as to generalize the generation to similar objects that the model was not explicitly trained on but had similar physical attributes. However, the action variants generated from DeepIMAGINATION have a lot of room to improve in different ways. There are definite problems with modal collapse to some extent in the network. This reduces its ability to generate action variants to match the 'true' distribution of appropriate action variants from the agent's action space. Current work on this aspect of the research focuses on improving the different architectural variants used to implement DeepIMAGINATION. Current research suggests that we could combine CVAE with adversarial approaches similar to [117] to arrive at better action variants. Future work could also add significant contributions to the field by exploring improvements in the interpretability of the model's latent space as well.

The physical attributes representation for objects that was used in the affordance-based action variant generation was iteratively refined by annotating Whose Line Is It Anyway? [175] Props game props and refining the schema as challenges were faced. The current representation conditions the DeepIMAGINATION model to implement affordance-based action generation. However, the current implementation of this vector representational

schema loses the spatial and ordering relationships between the respective parts of the prop. This could be addressed by using graph embeddings to learn the spatial and ordering encoding of the prop's parts while successfully being able to condition the DeepIMAGINATION model with only minimal adaptations to the generative model architecture. Therefore, this is a near term goal for the future of this representational schema in order to improve the affordance-based action variant generation.

A long-term goal for the current object representation would be to automatically derive the formal physical attributes for unfamiliar props based on their 3D models (either mesh-based or point cloud-based). This could be done by automatically segmenting the parts of the model and then classifying the segmented parts into their respective attributes (expanding on [231, 232, 233]). This addition remains a long-term goal for the research due to the significant computer vision challenges for developing a general system to perform this classification task.

The improvisational response strategies developed to optimize action space search and follow a creative arc in the agent's creative space were adapted from prior work in the LuminAI installation [50]. They were adapted from that research to work directly within the parameter space of the DeepIMAGINATION latent space. This was done by mapping strategies to conceptual 'moves' or operations within the latent space and utilizing the vector properties of that latent space. This has resulted in responsive creative arc negotiation behavior for the agent, that at least for observers and to different degrees for participants as well, elicits significant increases in perceptions of enjoyment, agent creativity, and coherence.

The formalized improvisational response strategies, however, do not cover a full spectrum of strategies that improvisers have been known to use [19, 58]. This includes both strategies that operate longer-term sequences than just the last action for example, being able to implement longer-term improvisational conventions about procedurally establishing and violating patterns. This particular example could be implemented using the creative arc

itself; however, this is certainly not guaranteed to be the case. Therefore, there is a larger question about how best to conceptualize or modulate interactions between a long-term action selection mechanism (creative arc negotiation) and the short-term opportunistic selection of actions (using improvisational response strategies). It is particularly important to balance the (relatively rare) potential experiential benefits of formalizing this knowledge with the amount of engineering and meta-authoring that the implementation process involves.

Another potential area of future work for adding to the improvisational response strategies is to directly attempt to generate actions with specific values in the agent's creative space. An example of this would be an unexpectedness-generation strategy, which would integrate more closely with the surprise calculation measures to directly generate actions that were unexpected. Another strategy in this vein would be a novelty-generation strategy. This is already almost possible using the similarity-based recall strategy since it can choose strategies at a given distance from the current action from the agent's action space. However, that is not a direct mapping to novelty, but a measure of gestural similarity (novelty is aggregated similarity compared across all the different aspects of a given action and bounded by a set of comparable actions). Quality-generation strategies would then need to integrate with and optimize the agent's quality metrics to generate actions with a given quality score directly. These purely theoretical examples of strategies might all be useful to the agent to speed up generation, however, at that point, the architecture potentially would not need the other strategies since the three of these strategies could directly perform creative arc negotiation in the agent's creative space.

Strategy selection to optimize the exploration of regions of the agent's action space is a component of CARNIVAL that has been reserved for future research. One possible method for accomplishing strategy selection toward this end would be to learn a policy for the relative change in the agent's position within the creative space, based on the selected strategy. The application of this learned policy would allow the agent to choose the top N

strategies that are most likely to move the agent in the direction of the target point on its current creative arc based on its current location in the creative space.

The CARNIVAL agent's use of creativity evaluation models is what enables it to follow a given creative arc. A few different techniques were used to implement evaluation models for the novelty, unexpectedness, and quality of perceived and generated actions. From sections 3.5.4 and 3.5.5, it can be seen that the improvisational agent was able to successfully influence the perceptions of observers and participants to significantly increase their perceptions of enjoyment (for observers), agent creativity, and coherence. Additionally, depending on the creative arc, (at least) observers were able to correctly identify some of the trends in the creative arc in terms of the properties of novelty, object surprise, action surprise, quality, and user-defined creativity. However, given the results of the paired comparison tasks that directly involved observers identifying which of two actions had a higher score of those same properties, the models performed badly. More detail is used to understand this finding in section 3.5.2, but from the current results, it is clear that the models need to be improved. Additional work needs to be done to mitigate the apparent paradox between the model's low prediction (or user recognition) accuracy across individual actions and the fact that heuristic-guided search using those same models seems to significantly positively impact observer perceptions of the resulting experience over longer-term comparisons of experience.

Future work to improve these models could include the following techniques. All three sets of models could use user feedback to tailor their recommendations to the individual or population of individuals over time. For the agent's model of unexpectedness, in particular, additional confidence-based modulation (or thresholding) as well as a more thorough computational affect model based on a validated theory of affect (such as appraisal theory [205] or the somatic marker hypothesis [206]) is required to develop the model into one that measures surprise (rather than unexpectedness). The model of unexpectedness could also use mechanisms for interactive learning to incorporate temporal expectations

across actions over time from the agent's growing experience (see [1] for some initial ideas). Alternatively, this temporally-sequential expectation could be learned using narrative knowledge-acquisition system such as [234, 235, 236]. The agent's quality evaluation model could also gain from incorporating additional quality heuristics for this domain. This set of heuristics could include the aesthetic pleasantness of an action, learned functions approximating user-generated ratings for actions, investigating computational models of humor, and adding other measures of action coherence over time.

The main hypothesis guiding the CARNIVAL architecture was that agents could mitigate the improvisational action selection problem to create experiences that were more enjoyable, with agents that seemed more creative, and to improvise performances that seemed more coherent by using creative arc negotiation for agent action selection. This approach seems to have been relatively successful, at least to observers of the installation, given the results of the creative arc identification and creative arc comparison (see sections 3.5.3 and 3.5.4). Initial results suggested that this was also the case, albeit to a lesser extent, for participants as well with the results from in-person preference studies (see section 3.5.5). This approach was originally conceived as being applicable not only to movement improv but also as a more general embodied model of improvisational creativity. With some adaptation, the core ideas of the model could also be distilled into a general model of improvisational creativity (regardless of the degree of embodiment). The validity of these claims for generality needs to be evaluated by adapting this model to other domains of embodied creativity. This is an important direction of future work for this research.

Another direction for improving this model of intrinsically motivated creative arc-negotiation for action space search could be to expand the number and kinds of spatial dimensions that the agent can measure and explore. These added dimensions could include measures of the social cognition like Guckelsberger, Salge, and Colton coupled empowerment maximization [160] or models of affect like [237]. However, there is a potential open question about the process of adding dimensions to the model. At what architectural level

would a new dimension be added to the model? Is it merely a measure of quality, or does it warrant addition as a fully explorable dimension in its own right? Additionally, until there is a mechanism for the agent to learn which arc to use or to personalize the arc to a particular individual, designers who create need to be able to map the intended experience to the dimensions of the arc. Finally, adding more dimensions for the agent to evaluate for every candidate action could also slow the agent down beyond a reasonable level of improvisational responsiveness.

3.8 Conclusion

This chapter presented my research into improvisational agents for performing movement improv in the Props game domain with non-experts within the Robot Improv Circus VR installation. The embodied virtual agents were controlled by the CARNIVAL architecture to enable them to perform creative arc negotiation for action selection in order to primarily address the improvisational action selection problem. The chapter discussed the details of the installation and the architecture before describing a number of validation and evaluation experiments used to investigate the claims in my thesis statement for this dissertation. The chapter concluded by discussing the limitations and future work of the research as well.

CHAPTER 4
CONCLUSION

4.1 Summary

My dissertation presented the following thesis statement about improvisational agents and embodied co-creativity for evaluation.

> Embodied agents that address the improvisational action selection problem using creative arc negotiation increase perceptions of enjoyment, agent creativity, and coherence in both observers and participants while performing movement improv with non-experts.

I described research into building and evaluating a movement improv installation between embodied improvisational agents and non-expert human participants as well as a human audience, in order to investigate my thesis statement. I described an interactive VR installation for playing the Props game with a virtual robot character, called the *Robot Improv Circus*, and the *Creative ARc Negotiating Improvisational Virtual Agent pLatform* (CARNIVAL) agent architecture, for enabling embodied virtual agents to improvise with non-expert human collaborators using creative arc negotiation as an action selection mechanism.

The improvisational action selection problem and how it could be addressed when situated within an improvisationally complex and semantically (or narratively) representational domain like the Props game (in comparison to prior work [50]) was investigated through this dissertation. The Props game involved representing and reasoning about gestural proto-narratives (as in prior work [50]) as well as about the objects in the agent's environment, and the relationship between affordance, object, and agent as a way to constrain as well as generalize action selection. Additionally, a novel form of intrinsically-motivated action

selection was developed called creative arc negotiation to tackle the improvisational action selection problem head-on in the more complex domain. Creative arc negotiation is action selection based on following a specified trajectory through a conceptual creative space of novelty, unexpectedness, and quality in concert with a fellow improviser's movements through that creative space as well. Operationalizing creative arc negotiation required the agent to evaluate perceived as well as generated actions through computational models of creativity evaluation as the novelty, unexpectedness, and quality of actions. The architecture also relied on affordance-based action variant generation and improvisational response strategies in order to perform real-time object-based gestural proto-narrative improvisation in the Props game with non-expert human collaborators, as required to investigate the claims in my thesis statement.

The CARNIVAL architecture and the Robot Improv Circus installation were evaluated using in-person user experience studies and observer rating studies to determine whether the claims made in my thesis statement were supported by evidence. These experiments (see section 3.5) demonstrate that the techniques used in the Robot Improv Circus successfully address the improvisational action selection problem and enable object-based gestural proto-narrative improvisation with non-expert human collaborators in the Props game. Analysis of the evidence produced by the evaluation studies showed that ultimately, it is **valid to conclusively state that embodied agents addressing the improvisational action selection problem using creative arc negotiation can perform movement improv with non-experts so that perceptions of agent creativity and coherence increase for both participants and audience members, but that perceptions of enjoyment only increase conclusively for observers. More study and data is required to show a conclusive increase in perceptions of enjoyment for participants of the installation.**

The evaluation of this research further illuminates where the system succeeds and where it needs to develop further in order to traverse the path towards unconstrained embodied narrative improvisation better. The approach used in CARNIVAL is very different from my

prior work in the LuminAI architecture [50], focusing on techniques from different parts of the artificial intelligence landscape. In so doing, CARNIVAL succeeds at addressing many issues that surround real-world, non-expert human-computer improvisation and embodied co-creativity, through the integration of multiple technical solutions and formalizations of ideas from human improvisational practice itself. In particular, this dissertation shows that CARNIVAL is demonstrably adept at generating and evaluating a variety of creative, coherent, and enjoyable actions over time and makes a definite impact on the experiences of performance audiences/observers and installation participants (though to a lesser degree).

4.2 Contributions

The contributions of my research in this dissertation are as follows.

- A model of affordance-based action variant generation for parameterized generation of action variants based on a given objects physical attributes.
- A formalized set of improvisational reasoning strategies for guiding an agents action space search based on previous experience and the current improvisational context.
- Computational models for evaluating the creativity of perceived and generated action variants in terms of their novelty, unexpectedness (as a measure of surprise), and quality (as a measure of value).
- A model of creative arc negotiation for improvisational action selection while performing movement improv with non-experts that increases both participant and observer perceptions of enjoyment, agent creativity, and coherence.
- A publicly disseminated and validated interactive installation where embodied agents can perform movement improv with non-experts.

4.3 Toward Unconstrained Embodied Narrative Improvisation

The research presented in this dissertation focuses on the knowledge representations, software architectures, and computational processes that can be used to develop improvisational agents for embodied co-creativity between non-expert humans and embodied virtual characters. My research on human-computer embodied improvisation in prior work with the LuminAI installation [50] and the research presented in this dissertation on the Robot Improv Circus (chapter 3) interactive installations form a body of work that focuses on addressing the various problems like the *improvisational action selection problem* that are inherent in human-computer movement improv. The direction of my research over time has formed a trajectory that points toward unconstrained embodied narrative improvisation in the future. In the following section, I present a sampling of significant remaining challenges and potential directions for continuing research toward unconstrained embodied narrative improvisation.

Unconstrained embodied narrative improvisation suffers from a severe knowledge-authoring bottleneck as shown by various previous human-agent improvisational systems [26, 56, 50, 51]. A developmental approach to interactively learning a larger subset of knowledge required for this improvisation through human-agent co-creativity in virtual environments was proposed in [1]. The proposed approach described a hierarchical (generalized directed hypergraphical) representation (see figure 4.1) in which the embodied knowledge of actions performed in the virtual world was perceived, segmented, clustered, interpreted, and abstracted into increasingly high-level knowledge that is temporally and causally sequenced into graphical structures in each layer of the representation over time. Note that this approach would require a large amount of human interaction for the agent to learn a realistically large space of knowledge for use in unconstrained embodied narrative improvisation with people. Therefore, the improvisational scenarios would need to be diverse and engaging enough to encourage continued participation.

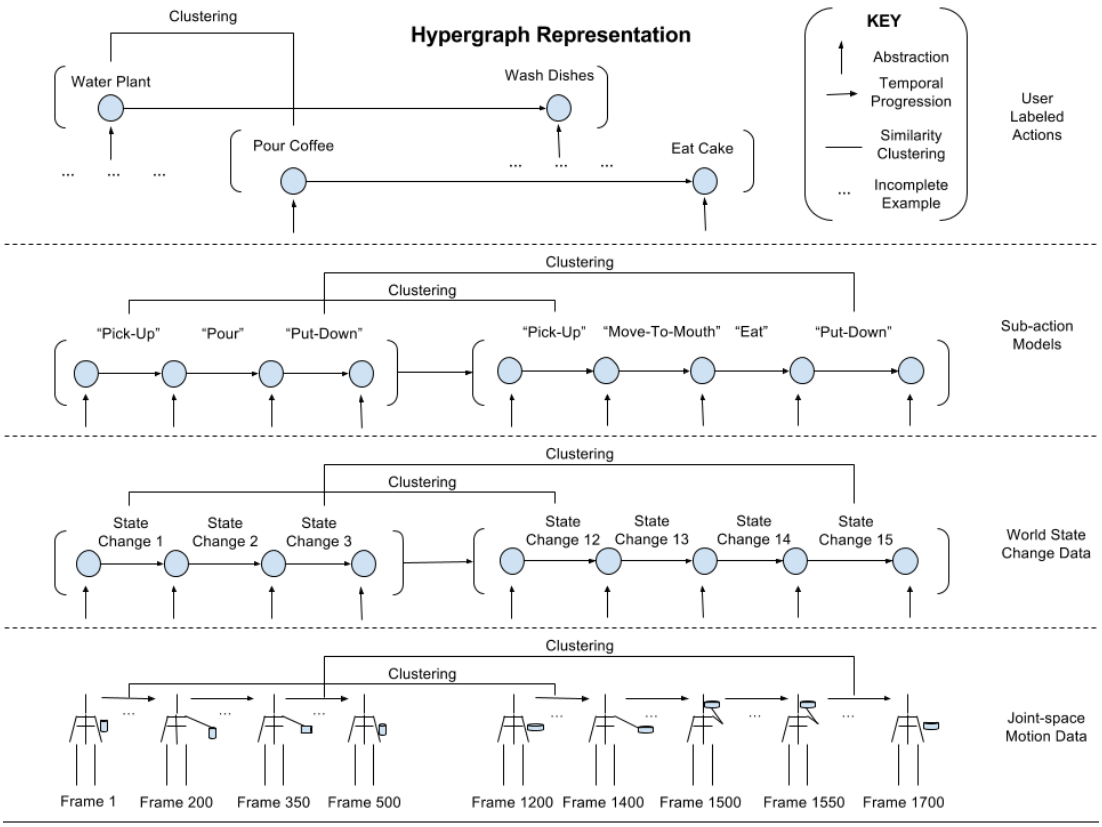

Figure 4.1: An example of learned knowledge represented within a proposed hypergraphical knowledge representation [1].

The hypergraphical knowledge representation proposed in [1] also presents a possible solution to another significant challenge in the development of improvisational agents for human-computer unconstrained embodied narrative improvisation — the deep integration of the embodied knowledge learned by the agent with high-level structured knowledge potentially obtained from mining corpora or from other large-scale knowledge repositories. Similarly, the hypergraphical knowledge representation could also serve as an integrative structure for fusing multimodal sensory percepts (speech, gesture, scene contents, images, etc.) and the resulting combined knowledge at different levels of abstraction. However, the proposed representation only demonstrated replay capabilities, with limited ability for transformation and variant generation. Given the useful generational properties of vector spaces and deep generative models as demonstrated within DeepIMAGINATION (see section 3.4.3), it would be useful to investigate how the hypergraphical representations can

be integrated together with vector spaces and deep generative models to enable knowledge learning and integration with powerful generative capabilities.

The knowledge-authoring bottleneck in unconstrained embodied narrative improvisation is exacerbated by the different kinds of knowledge that would be required for it to be successfully performed even with the use of interactive learning to mitigate the problem partially. A powerful addition to the interactive learning capabilities of an agent in this domain would be large-scale mining (that also integrates mined actions into existing learned knowledge) of required agent knowledge from the large number of videos on public repositories online (as briefly mentioned in section 3.7.1). My initial work in this area adapts [228] into an integrated pipeline for parsing the gestural content from fixed-camera YouTube videos of a single human figure performing a single action per video. Extensions to this initial work are necessary, in terms of automating action segmentation, annotating segmented actions with semantic content from various sources, learning from (or correctly handling) multiple figures in a single video, and compensating for video artifacts like camera motion. As research in visual scene understanding (such as in [238]) progresses beyond static, synthetic scenes, it will likely be possible to learn the accompanying 3D scene representations of the environments within which the sourced videos are situated. This would create a diverse set of virtual worlds for situating learned actions for the agent. It would also provide a diverse set of virtual worlds for situating new improvised scenarios with people as described above.

A significant focus of the CARNIVAL agent architecture is the creation of meaningful improvised experiences over time in ill-defined improvisational domains. This works for a domain like the Props game that has less narrative structure compared to unconstrained embodied narrative improvisation. Taking long-form improv theater as an example of the latter class of domain, it is likely that there will be strong expectations for coherent goal-driven behavior interspersed with unexpected or locally-incoherent behavior that is justified by subsequent actions (e.g., sequences of platform-tilt-justification [129] or finding the game

of the scene [132]). This requirement for fluidly navigating the spectrum of goal-directed to exploratory creative action generation is another significant challenge for improvisational agents performing unconstrained embodied narrative improvisation. The creative arc negotiation process offers a potential initial and partial solution to this problem as repeated rising and falling sections of a long, creative arc. However, creative arc negotiation is likely a complementary mechanism to the reasoning required rather than an exact solution to the previous problem.

The research presented in this dissertation is an initial exploration in the direction of unconstrained human-computer embodied narrative improvisation. None of the problems stated earlier in this section can easily be solved at this time. However, building on the initial results described in this dissertation and continuing future research along the directions described in this section, it is possible that unconstrained human-computer embodied narrative improvisation will someday number among the creative outlets, performing arts, interactive experiences, and expressive media where humans and computers can create seamlessly together.

REFERENCES

[1] M. Jacob, "Towards lifelong interactive learning for open-ended embodied narrative improvisation," in *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, ACM, 2017, pp. 502–507.

[2] M. Jacob and B. Magerko, "Interaction-based Authoring for Scalable Co-creative Agents," in *Proceedings of the Sixth International Conference on Computational Creativity (ICCC 2015)*, Provo, UT, 2015.

[3] N. Davis, M. Comerford, C.-P. Hsiao, M. Jacob, and B. Magerko, "An Enactive Characterization of Dyadic Pretend Play," in *Proceedings of the 10th ACM conference on Creativity and Cognition*, Glasgow: ACM, 2015.

[4] G. Hawthorne, E. M. Quintin, M. Saggar, N. Bott, E. Keinitz, N. Liu, Y. H. Chien, D. Hong, A. Royalty, and A. L. Reiss, "Impact and sustainability of creative capacity building: The cognitive, behavioral, and neural correlates of increasing creative capacity," in *Design thinking research*, Springer, 2014, pp. 65–77.

[5] B. Shneiderman, "Creativity support tools," *Communications of the ACM*, vol. 45, no. 10, pp. 116–120, 2002.

[6] B. D. Drake, G. M. Acosta, D. A. Wingard, and R. L. Smith, "Improving creativity, solving problems, and communicating with peers in engineering and science laboratories," *Journal of chemical education*, vol. 71, no. 7, p. 592, 1994.

[7] A. Newell, J. C. Shaw, and H. A. Simon, *The processes of creative thinking*. Rand Corporation Santa Monica, CA, 1959.

[8] M. A. Boden, *The creative mind: Myths and mechanisms*. Psychology Press, 2003, DOI: 10.1017/S0140525X0003569X, ISBN: 0203508521.

[9] S. Colton, "Creativity Versus the Perception of Creativity in Computational Systems," in *Proceedings of the AAAI Spring Symposium on Creative Systems*, 2008, pp. 14–20, ISBN: 9781577353591.

[10] S. Colton, J. W. Charnley, and A. Pease, "Computational creativity theory: The face and idea descriptive models.," in *Proceedings of the 2nd International Conference on Computational Creativity*, 2011, pp. 90–95.

[11] A. K. Jordanous, "Evaluating Computational Creativity: A Standardised Procedure for Evaluating Creative Systems and its Application," 2011.

[12] S. Colton, G. A. Wiggins, *et al.*, "Computational creativity: The final frontier?" In *Ecai*, Montpelier, vol. 12, 2012, pp. 21–26.

[13] N. M. Davis, "Creative sense-making: A cognitive framework for quantifying interaction dynamics in co-creation," PhD thesis, Georgia Institute of Technology, 2017.

[14] T. Lubart, "How can computers be partners in the creative process: Classification and commentary on the special issue," *International Journal of Human-Computer Studies*, vol. 63, no. 4-5, pp. 365–369, 2005.

[15] M. O. Riedl and B. ONeill, "Computer as audience: A strategy for artificial intelligence support of human creativity," in *Proc. CHI Workshop of Computational Creativity Support*, 2009.

[16] D. Long, M. Jacob, N. M. Davis, and B. Magerko, "Designing for Socially Interactive Systems," in *Proceedings of the Eleventh ACM Conference on Creativity and Cognition*, Singapore, 2017, p. 10.

[17] L. Winston and B. Magerko, "Turn-taking with improvisational co-creative agents," in *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2017.

[18] P. F. Berliner, *Thinking in jazz: The infinite art of improvisation*. University of Chicago Press, 2009.

[19] J. Pressing, "Cognitive processes in improvisation," in *Advances in Psychology*, vol. 19, Elsevier, 1984, pp. 345–363.

[20] R. K. Sawyer, "Improvisation," *Journal of Linguistic Anthropology*, vol. 9, no. 1-2, pp. 121–123, 1999.

[21] D. Mendona and W. A. Wallace, "A Cognitive Model of Improvisation in Emergency Management," *IEEE Transactions on Systems, Man and Cybernetics Part A: Systems and Humans*, vol. 37, no. 4, pp. 547–561, 2001.

[22] R. K. Sawyer and S. DeZutter, "Distributed creativity: How collective creations emerge from collaboration.," *Psychology of aesthetics, creativity, and the arts*, vol. 3, no. 2, p. 81, 2009.

[23] B. Magerko, W. Manzoul, M. Riedl, A. Baumer, D. Fuller, K. Luther, and C. Pearce, "An empirical study of cognition and theatrical improvisation," in *Proceedings of the seventh ACM conference on Creativity and cognition*, ACM, 2009, pp. 117–126.

[24] G. Hoffman and G. Weinberg, "Gesture-based human-robot jazz improvisation," in *Proceedings - IEEE International Conference on Robotics and Automation*, 2010, pp. 582–587, ISBN: 9781424450381.

[25] N. Davis, Y. Popova, I. Sysoev, C.-P. Hsiao, D. Zhang, and B. Magerko, "Building Artistic Computer Colleagues with an Enactive Model of Creativity," in *Proceedings of the Fifth International Conference on Computational Creativity (ICCC 2014)*, Ljubljana, Slovenia, 2014.

[26] B. O'Neill, A. Piplica, D. Fuller, and B. Magerko, "A knowledge-based framework for the collaborative improvisation of scene introductions," in *Proceedings of the 4th International Conference on Interactive Digital Storytelling*, vol. 7069 LNCS, Vancouver, Canada, 2011, pp. 85–96, ISBN: 9783642252884.

[27] R. Loughran and M. O'Neill, "Application domains considered in computational creativity.," in *ICCC*, 2017, pp. 197–204.

[28] R. J. Gerrig, *Experiencing narrative worlds: On the psychological activities of reading*. Yale University Press, 1993.

[29] M. Mateas and P. Sengers, *Narrative intelligence*.

[30] R. A. Mar, "The neuropsychology of narrative: Story comprehension, story production and their interrelation," *Neuropsychologia*, vol. 42, no. 10, pp. 1414–1434, 2004.

[31] H. P. Abbott, "The cambridge introduction to narrative. 2002," *Cambridge: Cambridge*, 2008.

[32] A. C. Graesser, K. Hauft-Smith, A. D. Cohen, and L. D. Pyles, "Advanced outlines, familiarity, and text genre on retention of prose," *The Journal of experimental education*, vol. 48, no. 4, pp. 281–290, 1980.

[33] A. C. Graesser, M. Singer, and T. Trabasso, "Constructing inferences during narrative text comprehension.," *Psychological review*, vol. 101, no. 3, p. 371, 1994.

[34] G. Lakoff and M. Johnson, *Metaphors we live by*. University of Chicago press, 2008.

[35] G. Prince, *Narratology: The form and functioning of narrative*. Walter de Gruyter, 2012, vol. 108.

[36] L.-C. Hydén, "Towards an embodied theory of narrative and storytelling," *The travelling concepts of narrative*, pp. 227–244, 2013.

[37] B. WIRE, *Worldwide Spending on Augmented and Virtual Reality Forecast to Reach $13.9 Billion in 2017, According to IDC*. 2017.

[38] FIVARS, "Festival of international virtual and augmented reality stories," *FIVARS*, 2017.

[39] N. Cooke and R. Stone, "RORSIM: A warship collision avoidance 3d simulation designed to complement existing Junior Warfare Officer training," *Virtual Reality*, vol. 17, no. 3, pp. 169–179, 2013.

[40] N. E. Seymour, A. G. Gallagher, S. A. Roman, M. K. Obrien, V. K. Bansal, D. K. Andersen, and R. M. Satava, "Virtual reality training improves operating room performance: Results of a randomized, double-blinded study," *Annals of surgery*, vol. 236, no. 4, pp. 458–464, 2002.

[41] S. Benford, J. Bowers, L. E. Fahlén, C. Greenhalgh, and D. Snowdon, "Embodiments, avatars, clones and agents for multi-user, multi-sensory virtual worlds," *Multimedia Systems*, vol. 5, no. 2, pp. 93–104, 1997.

[42] O. Labs, *Job Simulator: The 2050 Archives*. Owlchemy Labs, 2016.

[43] S. Studios, *Everest VR*. Slfar Studios, 2016.

[44] B. Magerko, "Measuring dramatic believability," *Intelligent Narrative Technologies*, pp. 79–82, 2007.

[45] 2019.

[46] 2019.

[47] R. Alves, M. Madeira, J. Ferrer, S. Costa, D. Lopes, B. M. da Silva, L. Sousa, J. Martins, and J. Rodrigues, "Fátima revisited: An interactive installation," *Proceedings of SGEM 2014*, pp. 141–148, 2014.

[48] L. MacDonald, J. Brosz, M. a. Nacenta, and S. Carpendale, "Designing the Unexpected: Endlessly Fascinating Interaction for Interactive Installations," in *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction - TEI '14*, 2015, pp. 41–48, ISBN: 9781450333054.

[49] L. A. Hernández-Ibáñez, V. Barneche-Naya, and R. Mihura-López, "Natural interaction and movement paradigms. a comparison of usability for a kinect enabled museum installation," in *International Conference on Learning and Collaboration Technologies*, Springer, 2016, pp. 145–155.

[50] M. Jacob, G. Coisne, A. Gupta, I. Sysoev, G. G. Verma, and B. Magerko, "Viewpoints AI," in *Proceedings of the Ninth Annual AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment (AIIDE)*, Boston, MA, 2013.

[51] B. Magerko, J. Permar, M. Jacob, M. Comerford, and J. Smith, "An Overview of Computational Co-creative Pretend Play with a Human," in *Proceedings of First Workshop on Playful Virtual Characters at the Fourteenth Annual Conference on Intelligent Virtual Agents*, Boston, MA, 2014.

[52] C. Mic, *La commedia dell'arte*. Verlag nicht ermittelbar, 1927.

[53] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[54] P. Abbeel and A. Y. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the twenty-first international conference on Machine learning*, ACM, 2004, p. 1.

[55] F. Torabi, G. Warnell, and P. Stone, "Behavioral cloning from observation," *ArXiv preprint arXiv:1805.01954*, 2018.

[56] A. Piplica, C. Deleon, and B. Magerko, "Full-body gesture interaction with improvisational narrative agents," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7502 LNAI, 2012, pp. 514–516.

[57] E. Şahin, M. Çakmak, M. R. Doğar, E. Uğur, and G. Üçoluk, "To afford or not to afford: A new formalization of affordances toward affordance-based robot control," *Adaptive Behavior*, vol. 15, no. 4, pp. 447–472, 2007.

[58] R. Hodson, *Interaction, improvisation, and interplay in jazz*. Routledge, 2007.

[59] L. Itti and P. Baldi, "Bayesian surprise attracts human attention," *Vision research*, vol. 49, no. 10, pp. 1295–1306, 2009.

[60] L. Macedo and A. Cardoso, "Modeling forms of surprise in an artificial agent," in *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 23, 2001.

[61] E. P. Torrance, *Torrance tests of creative thinking*. Princeton, N.J.: Personnel Press, Inc., 1968.

[62] H. G. Gough, "A creative personality scale for the adjective check list.," *Journal of personality and social psychology*, vol. 37, no. 8, p. 1398, 1979.

[63] D. M. Harrington, J. H. Block, and J. Block, "Testing aspects of carl rogers's theory of creative environments: Child-rearing antecedents of creative potential in young adolescents.," *Journal of personality and social psychology*, vol. 52, no. 4, p. 851, 1987.

[64] H. Gardner, *Creating minds: An anatomy of creativity seen through the lives of freud, einstein, picasso, stravinsky, eliot, graham, and ghandi*. Basic Civitas Books, 2011.

[65] J. R. Hayes, "Cognitive processes in creativity," in *Handbook of creativity*, Springer, 1989, pp. 135–145.

[66] R. K. Sawyer, "Group creativity: Musical performance and collaboration," *Psychology of music*, vol. 34, no. 2, pp. 148–165, 2006.

[67] R. Arnheim, *Visual thinking*. Univ of California Press, 1969.

[68] M. Sharples, "An account of writing as creative design," *The science of writing*, pp. 127–148, 1996.

[69] J. S. Gero, "Creativity, emergence and evolution in design," *Knowledge-Based Systems*, vol. 9, no. 7, pp. 435–448, 1996.

[70] Y. Chu and J. N. MacGregor, "Human performance on insight problem solving: A review," *The Journal of Problem Solving*, vol. 3, no. 2, p. 6, 2011.

[71] D. K. Simonton, *Creativity in science: Chance, logic, genius, and zeitgeist*. Cambridge University Press, 2004.

[72] R. Perez y Perez and M. Sharples, "MEXICA: A computer model of a cognitive account of creative writing," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 13, no. 2, pp. 119–139, 2001.

[73] B Daz-Agudo, P Gervs, and P. Gonzlez-Calero, "Poetry generation in COLIBRI," in *In Proceedings of the 6th European Conference on Advances in Case-Based Reasoning (ECCBR '02)*, 2002, pp. 73–87.

[74] J. L. Kolodner and D. B. Leake, "A Tutorial Introduction to Case-Based Reasoning," in *Case-Based Reasoning: EXPERIENCES, Lessons, and Future Directions*, D. B. Leake, Ed., 1996.

[75] S. Colton, "The Painting Fool: Stories from Building an Automated Painter," in *Computers and Creativity*, J. McCormack and M. dInverno, Eds., Springer Berlin Heidelberg, 2012, pp. 3–38, ISBN: 978-3-642-31726-2.

[76] S. Colton and G. a. Wiggins, "Computational creativity: The final frontier?" In *ECAI 2012: 20th European Conference on Artificial Intelligence: FRONTIERS in Artificial Intelligence and Applications*, L. De Raedt, C. Bessiere, and D. Dubois, Eds., vol. 242, DOI: 10.3233/978-1-61499-098-7-21, 2012, pp. 21–26, ISBN: 9781614990970.

[77] M. Cook, S. Colton, and J. Gow, "Automating game design in three dimensions," in *Proceedings of the AISB Symposium on AI and Games*, 2014, pp. 20–24.

[78] M. T. Pearce and G. A. Wiggins, "Evaluating cognitive models of musical composition," in *Proceedings of the 4th international joint workshop on computational creativity*, Goldsmiths, University of London, 2007, pp. 73–80.

[79] L. Candy and E. a. Edmonds, "Modeling co-creativity in art and technology," *Proceedings of the 4th conference on Creativity and Cognition*, pp. 134–141, 2002.

[80] N. Davis, C.-P. Hsiao, Y. Popova, and B. Magerko, "An enactive model of creativity for computational collaboration and co-creation," in *Creativity in the Digital Age*, Springer, 2015, pp. 109–133.

[81] P. Karimi, K. Grace, N. Davis, and M. L. Maher, "Creative sketching apprentice: Supporting conceptual shifts in sketch ideation," in *International Conference on-Design Computing and Cognition*, Springer, 2018, pp. 721–738.

[82] G. Smith, J. Whitehead, and M. Mateas, "Tanagra: Reactive planning and constraint solving for mixed-initiative level design," *IEEE Transactions on computational intelligence and AI in games*, vol. 3, no. 3, pp. 201–215, 2011.

[83] G. N. Yannakakis, A. Liapis, and C. Alexopoulos, "Mixed-initiative co-creativity.," in *FDG*, 2014.

[84] M. Nelson, S. Colton, E. Powley, S. Gaudl, P. Ivey, R. Saunders, B. Perez Ferrer, and M. Cook, "Mixed-initiative approaches to on-device mobile game design," 2016.

[85] M. Guzdial and M. Riedl, "An interaction framework for studying co-creative ai," *ArXiv preprint arXiv:1903.09709*, 2019.

[86] R. Swanson and A. S. Gordon, "Say anything: Using textual case-based reasoning to enable open-domain interactive storytelling," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 2, no. 3, p. 16, 2012.

[87] B. Samuel, M. Mateas, and N. Wardrip-Fruin, "The design of writing buddy: A mixed-initiative approach towards computational story collaboration," in *International Conference on Interactive Digital Storytelling*, Springer, 2016, pp. 388–396.

[88] R. Damiano, V. Lombardo, and A. Pizzo, "Doppiogioco. playing with the audience in an interactive storytelling platform," in *Conference on Complex, Intelligent, and Software Intensive Systems*, Springer, 2017, pp. 287–298.

[89] M. Roemmele and A. S. Gordon, "Automated assistance for creative writing with an rnn language model," in *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion*, ACM, 2018, p. 21.

[90] J. W. Davidson and N. Jordan, "Private Teaching, Private Learningi: An Exploration of Music Instrument Learning in the Private Studio, Junior and Senior Conservatories," in *International Handbook of Research in Arts Education*, Springer, 2007, pp. 729–754.

[91] T. W. Calvert, C. Welman, S. Gaudet, T. Schiphorst, and C. Lee, "Composition of multiple figure sequences for dance and animation," *The Visual Computer*, vol. 7, no. 2-3, pp. 114–121, Mar. 1991.

[92] K. Carlson, T. Schiphorst, and P. Pasquier, "Scuddle: Generating movement catalysts for computer-aided choreography.," in *ICCC*, 2011, pp. 123–128.

[93] S. F. Alaoui, K. Carlson, and T. Schiphorst, "Choreography as Mediated through Compositional Tools for Movement," *Proceedings of the 2014 International Workshop on Movement and Computing - MOCO '14*, pp. 1–6, 2014.

[94] K. Carlson, P. Pasquier, H. H. Tsang, J. Phillips, T. Schiphorst, and T. Calvert, "Cochoreo: A generative feature in idanceforms for creating novel keyframe animation for choreography," in *Proceedings of the Seventh International Conference on Computational Creativity*, 2016.

[95] L. Crnkovic-Friis and L. Crnkovic-Friis, "Generative choreography using deep learning," *ArXiv preprint arXiv:1605.06921*, 2016.

[96] J. A. Biles, "Genjam: Evolution of a jazz improviser," *Creative evolutionary systems*, vol. 168, p. 2, 2002.

[97] G. Hoffman and G Weinberg, "Shimon: An interactive improvisational robotic marimba player," *CHI'10 Extended Abstracts on Human Factors \ldots*, pp. 3097–3102, 2010.

[98] B. Thom, "Machine learning techniques for real-time improvisational solo trading.," in *ICMC*, 2001.

[99] R. M. Keller and D. R. Morrison, "A grammatical approach to automatic improvisation," in *Proceedings, Fourth Sound and Music Conference, Lefkada, Greece, July.*, 2007.

[100] D. J Mendona and W. A Wallace, "A cognitive model of improvisation in emergency management," *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 37, no. 4, pp. 547–561, 2007.

[101] B. Magerko, P. Dohogne, and C. DeLeon, "Employing Fuzzy Concepts for Digital Improvisational Theatre," *AIIDE*, 2011.

[102] A. Brisson, B. Magerko, and A. Paiva, "A computational model for finding the tilt in an improvised scene," in *International Conference on Interactive Digital Storytelling*, Springer, 2011, pp. 158–163.

[103] K. W. Mathewson and P. Mirowski, "Improvised theatre alongside artificial intelligences," in *Thirteenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2017.

[104] N. Davis, C.-P. Hsiao, K. Yashraj Singh, L. Li, and B. Magerko, "Empirically studying participatory sense-making in abstract drawing with a co-creative cognitive agent," in *Proceedings of the 21st International Conference on Intelligent User Interfaces*, ACM, 2016, pp. 196–207.

[105] D. Reidsma, A. Nijholt, R. Rienks, and H. Hondorp, "Interacting with a virtual rap dancer," in *Intelligent Technologies for Interactive Entertainment*, Springer Berlin Heidelberg, 2005, pp. 134–143, ISBN: 3540305092.

[106] L. J. Martin, B. Harrison, and M. O. Riedl, "Improvisational computational storytelling in open worlds," in *Interactive Storytelling: 9th International Conference on Interactive Digital Storytelling, ICIDS 2016, Los Angeles, CA, USA, November 15–18, 2016, Proceedings 9*, Springer, 2016, pp. 73–84.

[107] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.

[108] A. G. Billard, S. Calinon, and R. Dillmann, "Learning from humans," in *Springer Handbook of Robotics*, Springer, 2016, pp. 1995–2014.

[109] J. Saunders, C. L. Nehaniv, and K. Dautenhahn, "Teaching robots by moulding behavior and scaffolding the environment," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, ACM, 2006, pp. 118–125.

[110] T. Fitzgerald and A. Goel, "A case-based approach to imitation learning in robotic agents," in *Intl. Conf. on Case-Based Reasoning Workshop on Case-Based Agents*, 2014.

[111] V. Ng-Thow-Hing, P. Luo, and S. Y. Okita, "Synchronized gesture and speech production for humanoid robots," *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 4617–4624, 2010.

[112] T. S. K. Ikeuchi, "Synthesis of dance performance based on analyses of human motion and music," 2008.

[113] F. Ofli, E. Erzin, Y. Yemez, and A. M. Tekalp, "Learn2dance: Learning statistical music-to-dance mappings for choreography synthesis," *IEEE Transactions on Multimedia*, vol. 14, pp. 747–759, 2012.

[114] M. Mancini and G. Castellano, "Real-time analysis and synthesis of emotional gesture expressivity," 2007.

[115] M. Kipp, M. Neff, K. H. Kipp, and I. Albrecht, "Towards natural gesture synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis," in *IVA*, 2007.

[116] A. Augello, E. Cipolla, I. Infantino, A. Manfré, G. Pilato, and F. Vella, "Creative robot dance with variational encoder," *CoRR*, vol. abs/1707.01489, 2017. arXiv: 1707.01489.

[117] M. A. Kiasari, D. S. Moirangthem, and M. Lee, "Human action generation with generative adversarial networks," *CoRR*, vol. abs/1805.10416, 2018. arXiv: 1805.10416.

[118] L. Crnkovic-Friis and L. Crnkovic-Friis, "Generative choreography using deep learning," *CoRR*, vol. abs/1605.06921, 2016. arXiv: 1605.06921.

[119] T. Tang, J. Jia, and H. Mao, "Dance with melody: An lstm-autoencoder approach to music-oriented dance synthesis," in *Proceedings of the 26th ACM International Conference on Multimedia*, ser. MM '18, Seoul, Republic of Korea: ACM, 2018, pp. 1598–1606, ISBN: 978-1-4503-5665-7.

[120] D. Holden, J. Saito, and T. Komura, "A deep learning framework for character motion synthesis and editing," *ACM Trans. Graph.*, vol. 35, no. 4, 138:1–138:11, Jul. 2016.

[121] I. Habibie, D. Holden, J. Schwarz, J. Yearsley, and T. Komura, "A recurrent variational autoencoder for human motion synthesis," in *BMVC*, 2017.

[122] J. J. Gibson, "The senses considered as perceptual systems.," 1966.

[123] J. J. Gibson, *The ecological approach to visual perception*. Psychology Press, 1979.

[124] D. A. Norman, "The psychology of everyday things.," 1988.

[125] T. E. Horton, A. Chakraborty, and R. S. Amant, "Affordances for robots: A brief survey," *AVANT. Pismo Awangardy Filozoficzno-Naukowej*, vol. 2, pp. 70–84, 2012.

[126] L. Jamone, E. Ugur, A. Cangelosi, L. Fadiga, A. Bernardino, J. Piater, and J. Santos-Victor, "Affordances in psychology, neuroscience, and robotics: A survey," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 10, no. 1, pp. 4–25, 2016.

[127] T. A. Stoffregen, "Affordances as properties of the animal-environment system," *Ecological psychology*, vol. 15, no. 2, pp. 115–134, 2003.

[128] A. Chemero, "An outline of a theory of affordances," *Ecological psychology*, vol. 15, no. 2, pp. 181–195, 2003.

[129] K. Johnstone, *Impro: Improvisation and the theatre*. Theatre Arts Book, 1987.

[130] ——, *Impro for storytellers*. Routledge/Theatre Arts Books, 1999.

[131] 2019.

[132] M. Besser, I. Roberts, M. Walsh, J. Wengert, and D. Kantrowitz, *The upright citizens brigade comedy improvisation manual*. 2013.

[133] D. Fuller and B. Magerko, "Shared Mental Models in Improviational Theatre," in *Proceedings of 8th ACM Conference on Creativity and Cognition*, Atlanta, GA, 2011.

[134] A. Brisson, Magerko, Brian, and Paiva, Ana, "A Computational Model for Finding the Tilt in an Improvised Scene," Vancouver, canada: Springer, Nov. 2011.

[135] R. Hodhod, A. Piplica, and B. Magerko, "A formal architecture of shared mental models for computational improvisational agents," in *Intelligent Virtual Agents*, Springer, 2012, pp. 440–446, ISBN: 3642331963.

[136] M. Rhodes, "An analysis of creativity," *The Phi Delta Kappan*, vol. 42, no. 7, pp. 305–310, 1961.

[137] A. Jordanous, "Four pppperspectives on computational creativity in theory and in practice," *Connection Science*, vol. 28, no. 2, pp. 194–216, 2016.

[138] M. Kunda, K. McGreggor, and A. K. Goel, "A computational model for solving problems from the ravens progressive matrices intelligence test using iconic visual representations," *Cognitive Systems Research*, vol. 22, pp. 47–66, 2013.

[139] E. Policastro and H. Gardner, "11 from case studies to robust (generalizations: An approach to the study of creativity," *Handbook of creativity*, p. 213, 1999.

[140] C. Lamb, D. G. Brown, and C. L. Clarke, "Evaluating computational creativity: An interdisciplinary tutorial," *ACM Computing Surveys (CSUR)*, vol. 51, no. 2, p. 28, 2018.

[141] M. L. Maher, "Evaluating creativity in humans, computers, and collectively intelligent systems," in *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design*, Desire Network, 2010, pp. 22–28.

[142] G. Ritchie, "Some empirical criteria for attributing creativity to a computer program," *Minds and Machines*, vol. 17, no. 1, pp. 67–99, 2007.

[143] M. M. Perišić, M. Štorga, and J. Gero, "Situated novelty in computational creativity studies," in *10th International Conference on Computational Creativity ICCC19*, 2019.

[144] A. Pease, D. Winterstein, and S. Colton, "Evaluating machine creativity," in *Workshop on Creative Systems, 4th International Conference on Case Based Reasoning*, 2001, pp. 129–137.

[145] G. Fauconnier and M. Turner, *The Way We Think: Conceptual Blending and the Mind's Hidden Complexities*. Basic Books, 2003, ISBN: 0465087868, 9780465087860.

[146] D. Gentner and A. B. Markman, "Structure mapping in analogy and similarity.," *American psychologist*, vol. 52, no. 1, p. 45, 1997.

[147] G. Lakoff and M. Johnson, *Metaphors We Live By*. Chicago: The University of Chicago Press, 1980.

[148] M. Guzdial and M. Riedl, "Automated game design via conceptual expansion," in *Fourteenth Artificial Intelligence and Interactive Digital Entertainment Conference*, 2018.

[149] G. A. Wiggins, "Searching for computational creativity," *New Generation Computing*, vol. 24, no. 3, pp. 209–222, 2006.

[150] G. Wallas, *The Art of Thought*. Harcourt, Brace and Company, 1926.

[151] E. Sadler-Smith, "Wallas four-stage model of the creative process: More than meets the eye?" *Creativity Research Journal*, vol. 27, no. 4, pp. 342–352, 2015.

[152] M. Czikszentmihalyi, "Flow: The psychology of optimal experience," *Praha: Lidov Noviny*, 1996.

[153] S. Hélie and R. Sun, "Incubation, insight, and creative problem solving: A unified theory and a connectionist model.," *Psychological review*, vol. 117, no. 3, p. 994, 2010.

[154] R. A. Finke, T. B. Ward, and S. M. Smith, "Creative cognition: Theory, research, and applications," 1992.

[155] P. Johnson-Laird, "How Jazz Musicians Improvise," *Music Perception*, vol. 19, no. 3, pp. 415–442, 2002.

[156] D. E. Berlyne, "Curiosity and exploration," *Science*, vol. 153, no. 3731, pp. 25–33, 1966.

[157] K. Grace and M. L. Maher, "Specific curiosity as a cause and consequence of transformational creativity.," in *ICCC*, 2015, pp. 260–267.

[158] K. E. Merrick and M. L. Maher, *Motivated reinforcement learning: Curious characters for multiuser games*. Springer Science & Business Media, 2009.

[159] J. Schmidhuber, "Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts," *Connection Science*, vol. 18, no. 2, pp. 173–187, 2006.

[160] C. Guckelsberger, C. Salge, and S. Colton, "Intrinsically motivated general companion npcs via coupled empowerment maximisation," in *Computational Intelligence and Games (CIG), 2016 IEEE Conference on*, IEEE, 2016, pp. 1–8.

[161] C. Salge and C. Guckelsberger, "Does empowerment maximisation allow for enactive artificial agents?" In *Proceedings of the Artificial Life Conference 2016 13*, MIT Press, 2016, pp. 704–711.

[162] C. M. Fonseca and P. J. Fleming, "An overview of evolutionary algorithms in multiobjective optimization," *Evolutionary computation*, vol. 3, no. 1, pp. 1–16, 1995.

[163] J. Lehman and K. O. Stanley, "Abandoning objectives: Evolution through the search for novelty alone," *Evolutionary computation*, vol. 19, no. 2, pp. 189–223, 2011.

[164] D. Gravina, A. Liapis, and G. Yannakakis, "Surprise search: Beyond objectives and novelty," in *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference*, ACM, 2016, pp. 677–684.

[165] J.-B. Mouret, "Novelty-based multiobjectivization," in *New horizons in evolutionary robotics*, Springer, 2011, pp. 139–154.

[166] J. Bates, "Virtual reality, art, and entertainment," *Presence: Teleoperators & Virtual Environments*, vol. 1, no. 1, pp. 133–138, 1992.

[167] D. L. Roberts and C. L. Isbell, "A survey and qualitative analysis of recent advances in drama management," *International Transactions on Systems Science and Applications, Special Issue on Agent Based Systems for Human Learning*, vol. 4, no. 2, pp. 61–75, 2008.

[168] M. O. Riedl and V. Bulitko, "Interactive Narrative: An Intelligent Systems Approach.," *AI Magazine*, vol. 34, no. 1, pp. 67–77, 2013.

[169] M. Mateas and A. Stern, "Façade: An experiment in building a fully-realized interactive drama," in *Game developers conference*, vol. 2, 2003, pp. 4–8.

[170] J. Porteous, J. Teutenberg, D. Pizzi, and M. Cavazza, "Visual programming of plan dynamics using constraints and landmarks," in *Twenty-First International Conference on Automated Planning and Scheduling*, 2011.

[171] B. Magerko, "Evaluating preemptive story direction in the interactive drama architecture," *Journal of Game Development*, vol. 2, no. 3, pp. 25–52, 2007.

[172] M. O. Riedl, A. Stern, D. Dini, and J. Alderman, "Dynamic experience management in virtual worlds for entertainment, education, and training," *International Transactions on Systems Science and Applications, Special Issue on Agent Based Systems for Human Learning*, vol. 4, no. 2, pp. 23–42, 2008.

[173] M. Mateas and A. Stern, "Integrating plot, character and natural language processing in the interactive drama Faade," in *Proceedings of the 1st International Conference on Technologies for Interactive Digital Storytelling and Entertainment (TIDSE-03)*, 2003.

[174] M. O. Riedl and V. Bulitko, "Interactive narrative: An intelligent systems approach," *Ai Magazine*, vol. 34, no. 1, pp. 67–67, 2013.

[175] .

[176] M. Jacob and B. Magerko, "Creative arcs in improvised human-computer embodied performances," in *Proceedings of the 13th International Conference on the Foundations of Digital Games*, ACM, 2018, p. 62.

[177] N. Wouters, J. Downs, M. Harrop, T. Cox, E. Oliveira, S. Webber, F. Vetere, and A. Vande Moere, "Uncovering the honeypot effect: How audiences engage with public interactive systems," in *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, ACM, 2016, pp. 5–16.

[178] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using Part Affinity Fields," in *ArXiv preprint arXiv:1812.08008*, 2018.

[179] K. W. Mathewson and P. Mirowski, "Improbotics: Exploring the imitation game using machine intelligence in improvised theatre," *ArXiv preprint arXiv:1809.01807*, 2018.

[180] J. Pressing, "Psychological Constraints on Improvisation," in *In the Course of Performance: STUDIES in the World of Musical Improvisation*, B. Nettl and M. Russell, Eds., 1st ed., University Of Chicago Press, Dec. 1998, pp. 47–67, ISBN: 0-226-57411-3.

[181] W.-Y. Chan, H. Qu, and W.-H. Mak, "Visualizing the semantic structure in classical music works," *IEEE transactions on visualization and computer graphics*, vol. 16, no. 1, pp. 161–173, 2009.

[182] J. Braasch, "The $\mu\cdot$ cosm project: An introspective platform to study intelligent agents in the context of music ensemble improvisation," in *Sound-Perception-Performance*, Springer, 2013, pp. 257–270.

[183] E. H. Sparks, *Cantus firmus in mas and motet 1420-1520*. Univ of California Press, 1963.

[184] D. E. Berlyne, "Conflict, arousal, and curiosity.," 1960.

[185] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Advances in Neural Information Processing Systems*, 2015, pp. 3483–3491.

[186] C.-Y. Liou, W.-C. Cheng, J.-W. Liou, and D.-R. Liou, "Autoencoder for words," *Neurocomputing*, vol. 139, pp. 84–96, 2014.

[187] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *ArXiv preprint arXiv:1312.6114*, 2013.

[188] A. Lenci, "Distributional semantics in linguistic and cognitive research," *Italian journal of linguistics*, vol. 20, no. 1, pp. 1–31, 2008.

[189] K. M. Varadarajan and M. Vincze, "Afnet: The affordance network," in *Asian Conference on Computer Vision*, Springer, 2012, pp. 512–523.

[190] D. Song, K. Huebner, V. Kyrki, and D. Kragic, "Learning task constraints for robot grasping using graphical models," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2010, pp. 1579–1585.

[191] A. Stoytchev, "Behavior-grounded representation of tool affordances," in *Proceedings of the 2005 ieee international conference on robotics and automation*, IEEE, 2005, pp. 3060–3065.

[192] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: From sensory–motor coordination to imitation," *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 15–26, 2008.

[193] K. M. Varadarajan and M. Vincze, "Semantic saliency using k-tr theory of visual perception," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, IEEE, 2012, pp. 3676–3679.

[194] M. Jacob, A. Zook, and B. Magerko, "Viewpoints ai: Procedurally representing and reasoning about gestures.," in *DiGRA conference*, 2013.

[195] A. Guttman, *R-trees: A dynamic index structure for spatial searching*, 2. ACM, 1984, vol. 14.

[196] J. Altarriba, L. M. Bauer, and C. Benvenuto, "Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words," *Behavior Research Methods, Instruments, & Computers*, vol. 31, no. 4, pp. 578–602, 1999.

[197] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *ArXiv preprint arXiv:1607.01759*, 2016.

[198] L. Van Der Maaten, "Learning a parametric embedding by preserving local structure," in *Artificial Intelligence and Statistics*, 2009, pp. 384–391.

[199] M. Jacob, P. Chawla, L. Douglas, Z. He, J. Lee, T. Sawant, and B. Magerko, "Affordance-based generation of pretend object interaction variants for human-computer improvisational theater,"

[200] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, *TensorFlow: Large-scale machine learning on heterogeneous systems*, Software available from tensorflow.org, 2015.

[201] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014. eprint: `1412.6980`.

[202] S. Kullback, *Information theory and statistics*. Courier Corporation, 1997.

[203] A. Roberts, J. Engel, C. Raffel, C. Hawthorne, and D. Eck, "A hierarchical latent vector model for learning long-term structure in music," in *ICML*, 2018.

[204] K. Grace, M. L. Maher, D. Fisher, and K. Brady, "Data-intensive evaluation of design creativity using novelty, value, and surprise," *International Journal of Design Creativity and Innovation*, vol. 3, no. 3-4, pp. 125–147, 2015.

[205] K. R. Scherer, "Appraisal theory," *Handbook of cognition and emotion*, pp. 637–663, 1999.

[206] A. R. Damasio, "The somatic marker hypothesis and the possible functions of the prefrontal cortex," *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 351, no. 1346, pp. 1413–1420, 1996.

[207] A. Jordanous, "Creativity vs quality: Why the distinction matters when evaluating computational creativity systems," AISB, 2018.

[208] M. L. Maher, "Evaluating creativity in humans, computers, and collectively intelligent systems," in *Proceedings of the 1st DESIRE Network Conference on Creativity and Innovation in Design*, Desire Network, 2010, pp. 22–28.

[209] M. A Boden, *The creative mind: Myths and mechanisms*. Routledge, 2004.

[210] R. L. Thorndike, "Who belongs in the family?" *Psychometrika*, vol. 18, no. 4, pp. 267–276, 1953.

[211] L. Van Der Maaten, "Accelerating t-sne using tree-based algorithms," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 3221–3245, 2014.

[212] K. Grace and M. L. Maher, "Surprise and reformulation as meta-cognitive processes in creative design," in *Proceedings of the third annual conference on advances in cognitive systems ACS*, 2015, p. 8.

[213] ——, "What to expect when you're expecting: The role of unexpectedness in computationally evaluating creativity.," in *ICCC*, 2014, pp. 120–128.

[214] J. Sprott, "Some simple chaotic jerk functions," *American Journal of Physics*, vol. 65, no. 6, pp. 537–543, 1997.

[215] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *ArXiv preprint arXiv:1511.06349*, 2015.

[216] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[217] W. H. Kruskal and W. A. Wallis, "Use of ranks in one-criterion variance analysis," *Journal of the American statistical Association*, vol. 47, no. 260, pp. 583–621, 1952.

[218] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics bulletin*, vol. 1, no. 6, pp. 80–83, 1945.

[219] K. Grace, M. L. Maher, M. Mohseni, and R. P. y Pérez, "Encouraging p-creative behaviour with computational curiosity.," in *ICCC*, 2017, pp. 120–127.

[220] D. Long, M. Jacob, and B. Magerko, "Designing co-creative ai for public spaces," in *Proceedings of the 2019 on Creativity and Cognition*, ACM, 2019, pp. 271–284.

[221] Aristotle. and S. H Butcher, *Poetics*. Hill and Wang, 1969.

[222] G. Freytag, *Freytag's technique of the drama: An exposition of dramatic composition and art*. Scholarly Press, 1896.

[223] K. Vonnegut, "At the blackboard," *Lapham's Quaterly*, 2005.

[224] A. J. Reagan, L. Mitchell, D. Kiley, C. M. Danforth, and P. S. Dodds, "The emotional arcs of stories are dominated by six basic shapes," *EPJ Data Science*, vol. 5, no. 1, p. 31, 2016.

[225] M. T Kelso, P. Weyhrauch, and J. Bates, "Dramatic presence," *Presence: The Journal of Teleoperators and Virtual Environments*, vol. 2, no. 1, pp. 1–15, 1993.

[226] B. Laurel, "Computers as theatre reading," *Mas: Addison-Wesley Publishing Company*, 1991.

[227] M. Guzdial and M. O. Riedl, "Combinets: Creativity via recombination of neural networks," *ArXiv preprint arXiv:1802.03605*, 2018.

[228] D. Pavllo, C. Feichtenhofer, D. Grangier, and M. Auli, "3d human pose estimation in video with temporal convolutions and semi-supervised training," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7753–7762.

[229] W.-C. Hu, C.-H. Chen, T.-Y. Chen, D.-Y. Huang, and Z.-C. Wu, "Moving object detection and tracking from video captured by moving camera," *Journal of Visual Communication and Image Representation*, vol. 30, pp. 164–180, 2015.

[230] L. Ponzanelli, G. Bavota, A. Mocci, M. Di Penta, R. Oliveto, B. Russo, S. Haiduc, and M. Lanza, "Codetube: Extracting relevant fragments from software develop-

ment video tutorials," in *Proceedings of the 38th International Conference on Software Engineering Companion*, ACM, 2016, pp. 645–648.

[231] M. Schoeler and F. Wörgötter, "Bootstrapping the semantics of tools: Affordance analysis of real world objects on a per-part basis," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 8, no. 2, pp. 84–98, 2015.

[232] P. Abelha and F. Guerin, "Transfer of tool affordance and manipulation cues with 3d vision data," *ArXiv preprint arXiv:1710.04970*, 2017.

[233] W. Wu, Z. Qi, and L. Fuxin, "Pointconv: Deep convolutional networks on 3d point clouds," *ArXiv preprint arXiv:1811.07246*, 2018.

[234] J. Orkin and D. Roy, "The restaurant game: Learning social behavior and language from thousands of players online," *Journal of Game Development*, vol. 3, no. 1, pp. 39–60, 2007.

[235] B. Li, S. Lee-Urban, and M. Riedl, "Crowdsourcing interactive fiction games.," in *FDG*, Citeseer, 2013, pp. 431–432.

[236] K. Pichotta and R. Mooney, "Statistical script learning with multi-argument events," in *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, 2014, pp. 220–229.

[237] A. Mehrabian, "Analysis of the big-five personality factors in terms of the pad temperament model," *Australian journal of Psychology*, vol. 48, no. 2, pp. 86–92, 1996.

[238] S. A. Eslami, D. J. Rezende, F. Besse, F. Viola, A. S. Morcos, M. Garnelo, A. Ruderman, A. A. Rusu, I. Danihelka, K. Gregor, *et al.*, "Neural scene representation and rendering," *Science*, vol. 360, no. 6394, pp. 1204–1210, 2018.