STARS

University of Central Florida

STARS

Electronic Theses and Dissertations, 2004-2019

2005

Implementing Lexical And Creative Intentionality In Synthetic Personality

Erik Vick
University of Central Florida

Part of the Other Psychology Commons

Find similar works at: https://stars.library.ucf.edu/etd

University of Central Florida Libraries http://library.ucf.edu

STARS Citation

University of Central Florida

STARS
Showcase of Text, Archives, Research & Scholarship

IMPLEMENTING LEXICAL AND CREATIVE INTENTIONALITY IN SYNTHETIC
PERSONALITY

by

ERIK HENRY VICK
B.A. University of Central Florida, 1992
M.S.C.S University of Central Florida, 1998

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Modeling and Simulation
in the College of Arts and Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2005

ABSTRACT

Creating engaging, interactive, and immersive synthetic characters is a difficult task and evaluating the success of a synthetic character is often even more difficult. The later problem is solved by extending Turing's Imitation Game thusly: computational construct should be evaluated based on the criteria of how well the character can mimic a human. In order to accomplish a successful evaluation of the proposed metric, synthetic characters must be consistently believable and capable of role-appropriate emotional expression. The author believes traditional synthetic characters must be improved to meet this goal.

For a synthetic character to be believable, human users must be able to perceive a link between the mental state of the character and its behaviors. That is to say, synthetic characters must possess intentionality. In addition to intentionality, the mental state of the character must be human-like in order to provide an adequate frame of reference for the human users' internal simulations, to wit, the character's mental state must be comprised of a synthetic model of personality, of personality dynamics, and of cognition, each of which must be psychologically valid and of sufficient fidelity for the type of character represented. The author proposes that synthetic characters possessing these three models are more accurately described as synthetic personalities.

The author proposes and implements computational models of personality, personality dynamics, and cognition in order to evaluate the psychological veracity of these models and computational equivalence between the models and the implementation as a first step in the process of creating believable synthetic personalities.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE: INTRODUCTION

Since the beginnings of electronic computation, scientists have been seeking an answer to the question: "Can machines think?" In 1950, Alan Turing imagined a way to answer this difficult question (Turing, 1950). Turing proposed an "imitation game", in which there are three participants, two humans and the machine to be tested. One human plays the role of an interrogator or judge, while the other human and the machine intelligence play the role of respondents. In this game, the question "Can machines think?" may be answered in the positive if the interrogator is unable to determine which of the two respondents is human. This introduces an important concept: computational constructs should be evaluated based on the criteria of how well these constructs can mimic a human.

Over fifty years later, a new form of computational construct that is consistent with Turing's notions has emerged in an effort to create more realistic avatars in digital media. Generally referred to as virtual or synthetic characters, these are computational characters with at least enough autonomy to react to situations or scenarios in a manner that is engaging to a human participant and that supports the suspension of disbelief.

Believable, engaging characters have the ability to capture our attention and leave a lasting impression (Rickel, 2001). A character is defined by WordNet as "an actor's portrayal of someone in a play" whereas the American Heritage Dictionary defines a character as "a person portrayed in an artistic piece, such as a drama or a novel." Therefore, a synthetic character should be defined as a virtual portrayal of some person or virtual entity, as opposed to a synthetic person. By this definition, synthetic characters are not required to have emotions, intellect, or

human-like traits, but only be able to emulate these characteristics. That is, if the characters themselves, and the character's emulation of these human-like traits are believable.

One of the strongest methods of communication, teaching, and human interaction is the art of story telling. A sense of story (i.e., the elements of plot and characterization) is crucial in the creation of engagement and immersion (Douglas and Hargadon, 2000). In this regard, well-formed synthetic characters can provide this sense of story and harnesses the power of 'make believe' in digital media, by giving believable and occasionally surprising performances (Rousseau and Hayes-Roth, 1998). Further, these characters can be used in interactive story telling, can be implemented as a part of intelligent tutoring systems for education and training, enrich game play in video games, populate synthetic environments, and to develop intelligent interfaces between humans and computers.

Synthetic characters offer the potential for a variety of distinct new capabilities to digital media. These capabilities include:

1. Interactive Demonstration

2. Navigational Guidance

3. Gaze and Gesture

4. Nonverbal feedback

5. Conversational Signals

6. Emotions and Personality

7. Virtual Teammates

8. Story and Character (Rickel, 2001)

These capabilities can be exploited to further the level of immersion and engagement felt by the human user. Some characters may motivate the human participant by conveying excitement and

2

enthusiasm, some by providing familiar nonverbal and verbal signs of encouragement, and some by building on the strengths of fictional character development to create a sense of interest and leave a memorable impression.

Yoon et al. (2000) states: "Human participants can comfortably interact with interactive characters only when they feel they can understand what is going on in the character's mind" and that "the character's actions must be driven by its beliefs and desires." These statements address a concept known as intentionality, that is, the perceived link between a character's action and its beliefs and desires. In order for this link to be perceived and believed, the character must be able to communicate appropriately and consistently express emotional and social cues (Rousseau and Hayes-Roth, 1998).

In general terms, communication between humans may be either lexical or creative in nature (Poggi, 2001). A lexical system is one that relies on a lexicon of syntax-semantics pairs, while a creative system is one that relies on a set of encoded rules to create new syntax for a given meaning or to decipher a meaning from a given syntax (Poggi, 2001). Any lexical system and any creative system developed for a synthetic character must operate within the scope of the established role of the character and must be appropriate for the situation.

For the purposes of the present work, let lexical intentionality be defined as "expressing the internal beliefs and desires of a synthetic character either through a lexicon of natural language or through a lexicon of gesture, gaze, and expression."

In natural language, the basic ontology of words and meanings, and the parts of words and their meanings are "lexical atoms" (Poggi, 2001). Equipping a synthetic character with lexical verbal skills is necessary, but clearly not sufficient. Without some means of establishing

the context of the lexical atom, in many cases it is impossible to distinguish its semantics, beyond iconic, innocuous phrases.

For the purposes of the present work, let creative intentionality be defined as "expressing the internal beliefs and desires of a synthetic character either through a set of rules that allows the creation of syntax-semantics pairs for natural language, gestures, gaze, expression, or any combination of these."

With respect to natural language, rules of composition are representative of creative communication (Poggi, 2001). These rules give us the ability to create new words from atomic parts, sentences from lexical atoms, and corpora from sentences. A creative system for verbal communication is built upon the lexical system of the language used. Given that, the creative system is necessary and sufficient to allow creative intentionality in synthetic characters.

It should be noted that combining any lexical atoms, even two atoms from disparate lexicons, in accordance with a creative rule set constitutes creative intentionality. For example, humans routinely combine gestures and expressions to create new meanings. Take the case of waving. What is communicated by waving while smiling is certainly different than what is communicated by waving while frowning. The ability to create compound communications by concatenating gesture adds depth to the synthetic character's ability for expression, and therefore increases the intentionality of the character.

Creating a synthetic character that displays lexical and creative intentionality is a difficult and complex task. This goal cannot be accomplished by brute force since the interactions between the human user and the synthetic character will be too complex to reliably pre-script all the possible exchanges (Yoon et al., 2000). Any synthetic character designed to successfully mimic a human must implement a cognitive model of sufficient depth to allow creative

4

communication in order to support lexical and creative intentionality. However, cognitive models alone are insufficient.

Reilly and Bates (1995) identify personality as the most important feature to add to a synthetic character in order to foster believability. Further, interesting personalities heighten the immersion and engagement of human participants in traditional media (Rousseau and Hayes-Roth, 1998). Research shows that people are willing to respond to machines that are endowed with human-like personality characteristics as if the machines actually have personalities (Reeves and Nash, 1996), hence by adding personality to synthetic characters, immersion and engagement with these characters is increased.

Rousseau and Hayes-Roth (1998) opine that personality traits "correspond to patterns of behavior and modes of thinking that determine a person's adjustment to the environment." WordNet defines personality as the "complex of all the attributes - behavioral, temperamental, emotional, mental - that characterize a unique individual." Similarly, Webster's Revised Unabridged dictionary defines personality as "that which constitutes distinction of person – individualness". The American Heritage Dictionary offers that personality is "the distinctive traits of a person." Hence, a synthetic personality may be defined as a composite of characteristics that make the entity unique. Certainly, beliefs, desires, emotions, and reactions to social cues, are part of the characteristics that create individuality. Therefore, personality may be seen as the over-riding force that dictates a synthetic personality's actions.

The author's primary focus is on creating synthetic characters powerful enough to immerse and engage a human user, and versatile enough to work across multiple problem domains and respond appropriately to novel situations. In other words, the author seeks to create a synthetic character capable of lexical and creative intentionality, i.e., a machine that is able to

appear to be human to its human users. The author believes that in order to create a compelling synthetic character that is able to consistently function in a human manner; it must be able to go further than appearing to have personality, intention, intellect, and emotion. It must have them.

Characteristics of Effective Synthetic Characters

To be effective, synthetic characters must be believable, comfortable, and capable of real time interaction (Kille and Warwick, 2001). They should be adaptable and capable of responding to user's cognitive styles, implicit preferences, and explicit responses, as well. In order to create the highest levels of realism, engagement, and immersion, synthetic characters must be able to model intellect. Intelligence is defined in the American Heritage Dictionary as the "faculty of thought and reason, or the ability to acquire and apply knowledge." Similarly, Webster's Revised Dictionary defines intelligence as "the capacity to know or understand." By combining these definitions, intelligence may be defined in part by the ability to cogitate. The author uses the following definition of cognition: "the collection of mental processes and activities used in perceiving, learning, remembering, thinking, and understanding, and the act of using these processes" (Ashcraft, 1998). Therefore, the artificial intelligence component of any synthetic character must contain a suite of functions that allow the character to perceive, learn, remember, think, and understand.

Recent discoveries in the fields of Cognitive Science, Psychology and Neurology offer support to the importance of emotion to human intelligence (El Nasir et al., 1999; Chen et al., 2001 a). Hence, these characters must be able to model appropriate emotive components that are applicable to the character (El Nasir et al., 1999; Chen et al., 2001 a; Gratch and Marsella, 2001).

As mentioned previously, synthetic characters must be able to convey intentionality and communicate effectively. Since the character will be communicating with a human, and therefore must be able to function in an incomplete and unpredictable problem domain, the character must be highly adaptable and robust (Yoon et al., 2000). However, it is not sufficient for the synthetic character to be adaptable; the adaptation must be believable (Yoon et al., 2000). That is to say, the adaptation to user input must be shown in incremental and appropriate responses to the user's attempts to impact the interaction.

Fundamentally, a synthetic character must be more than a pre-scripted character, since is it very difficult for a designer to predict all possible scenarios and develop appropriate responses for them (Yoon et al., 2000). At minimum, it must be a reactive system, that is, a system in which no prediction takes place, but rather the system waits and processes user inputs without context. However, those synthetic characters that are able to perform anticipatory planning could be the most effective. In an anticipatory system, the synthetic character uses prediction to narrow the scope of the problem domain and to establish frames of reference to proactively eliminate possible interpretations that do not make sense contextually.

In summary, to be effective in its role, any synthetic character must:

1. be believable
2. be capable of real-time interaction
3. make the human user feel comfortable interacting with it
4. respond to the user's cognitive style and react appropriately to the user's implicit and explicit preferences
5. be adaptable
6. learn

7. have persistent memory

8. model emotion

9. communicate effectively

10. anticipate the user.

Common Roles Of Synthetic Characters

Synthetic characters enrich human experience with digital media. They can be used to tell stories interactively, develop intelligent tutoring systems for education and training, enrich the engagement in video games, populate synthetic environments, and to develop intelligent interfaces between humans and computers.

Generally speaking, synthetic characters should be able to demonstrate tasks in a virtual environment while providing commentary describing the task. By using adaptive planning, the character may also respond to interruptions from the user, and tailor its demonstration to the level of the user; thereby spending the most time on subjects about which the user has the most questions.

Synthetic characters also play a role in navigation through the simulation or information space. While navigation is generally construed to mean navigation of a virtual, three-dimensional space that has a real-world counterpart, it applies to navigation of computer systems as well. Not only do synthetic agents lead users to functionality they desire (e.g., by navigating menus, directory structures, etc.), but also allow for faster and more targeted traversal of informational spaces.

For example, in intelligent tutoring systems that do not use embodied synthetic characters, arrows and highlight colors are frequently used to direct a student's attention. Embodied synthetic characters, however, can direct the user's attention with more natural seeming gestures or by directing its gaze at the item in question

Humans use nonverbal feedback cues to influence one another in every interaction. Facial expressions, body language, gaze, and gestures are used in concert to express opinions and communicate level of engagement.

For example, characters may smile and nod to convey praise or agreement, look puzzled to express confusion, or employ virtual clowning to show excitement and enthusiasm and to further engage the user in the interface. By providing appropriate levels of emphasis, synthetic characters will be viewed as more life-like. Also, conversational signals used by the character will assist in regulating turn-taking and will provide back-channel feedback to the human user. Although not within the scope this dissertation, it is important that synthetic characters are also endowed with this capability and related research in this area will be discussed.

By presenting the user with an interface that they believe to have identifiable personality and emotions, engagement with the system may be increased. Additionally, problems that might prove frustrating to the user in a traditional interface may be cast in a humorous light by a character's personality.

Since humans are social creatures and often motivated and informed by social cues, synthetic characters may also be used to bolster the efficacy of video games, simulations, and other forms of interactive media by providing us with virtual teammates, virtual opponents, virtual mentors and coaches, guides, by sharing ownership of tasks and duties, and by offering encouragement, an enthusiastic ear, and a supportive environment, etc. Level of effort can be

9

elevated and creativity can be sparked by a sense of play, interaction, and companionship.

Additionally, leadership skills may be trained and polished using a system of synthetic

teammates endowed with human-like personalities.

CHAPTER TWO: FROM ACTORS TO CHARACTERS TO PERSONALITIES

A plethora of research into synthetic characters is ongoing in the academic sector, the private sector, and in government-sponsored programs, such as that conducted by Bates, 1994; Blumberg, 1995; Chen et al., 2001a; Chen et al., 2001b; El Nassir, 1999; Gratch and Marsella, 2001; Hayes-Roth, 1997; Hibbard, 2002; Kshirsagar, 2002; Paiva et al., 2001; Rickel, 2001; Rousseau and Hayes-Roth, 1998, Yoon et al., 2000, and many, many others. This research varies in the degree of importance placed on creating a construct that actually functions as humans do, versus a construct that appears to function as humans do.

The field of Artificial Intelligence has struggled to find an adequate definition of human intelligence and "the human mind". For many years, researchers tried to model human intelligence by modeling the logical brain in isolation. Additionally, research has become increasingly granular, often targeting a specific mental function or suite of functionality, to the exclusion of the concept of "the mind" as a whole. This research is certainly valuable, however, and is used by the author as a suite of skills or a set of brushes that might be used by a painter to create an independent work.

Chen et al. treats the mind as a control system and takes a computational approach to model mental processes (Chen et al., 2001 a). The implemented model of the mind incorporates a motivational system and an emotion system. These two systems interact so that motivations affect the current emotional state and so that emotions affect the motivational state.

The motivation system evaluates current motivations, activates the motivation that is most appropriate given the situation, chooses the active goal, and feeds goals to the emotion system. It relies on a static hierarchy of motivations that is subsumed by a dynamic component.

11

The emotion system is tasked with generating affective responses to the environment with respect to the agent's goals and motivations. The emotion generated by the system influences goal generation, and effects the selection of action.

In Gratch and Marsella (2001), the problem of modeling emotive influence is examined in two complimentary ways, one developed by Gratch and one developed by Marsella. Gratch's Emile system focuses on how the agent's perception of the environment impacts its emotional state, while Marsella's IPD system focuses on the influences of emotion, cognition, and behavior have on each other (Gratch and Marsella, 2001).

Gratch's approach suggests that by maintaining an explicit plan, complex and powerful emotions such as anger, hope, fear, and anticipation are possible. This method also offers dynamic adaptation of emotions by enabling appraisals of the current state. It does not, however, address how this dynamic emotive state affects the agent's behavior (Gratch and Marsella, 2001).

Marsella's work organizes behaviors into five distinct focus modes. These modes are specific lexicons of behavior that are tightly coupled with emotional states. In fact, transition between modes is governed by the current emotional state of the agent. These focus modes are designed to heighten the ability of the agent to convey intentionality (Gratch and Marsella, 2001).

When both Gratch's and Marsella's work are combined, the result is distinct behavioral focus modes that change due to a dynamic emotional state based on the appraisal of the current environment and the agent's plan. This is extended to allow behavior to influence the cognition of the agent. The focus mode is tied to a domain specific stimuli ranking mechanism and filter set that govern the agent's sensitivity to external influences (Gratch and Marsella, 2001).

In El Nasir et al. (1999), the generation of emotions is tied closely to the process of learning. It is argued that the process of event prediction, a direct precursor of emotion generation in this model, is impacted by information learned about specific events and about specific actions of users. Learning about pleasing and displeasing actions alters the internal model of the user, a precursor of behavior selection. Finally, Pavlovian stimulus-response conditioning may directly modify event-emotion association, also a precursor of emotion generation (El Nasir et al., 1999).

Rousseau and Hayes-Roth (1998) constructed a synthetic actor endowed with a model of personality based on social learning theory. Rousseau and Hayes-Roth implemented this model by distinguishing between personality traits, mood, and emotion. Although they state that "personality is a persistent characteristic that changes little and slowly, if at all, over time", Rousseau and Hayes-Roth point out that the strength of the "social-psychology model" of personality is that it allows the actors' personality traits to change as a result of learning from each interaction (Rousseau and Hayes-Roth, 1998). Each general behavior is specified in advance, allowing for conditions on personality traits to be tested. The implementation of mood is similar, as is the implementation of emotion (called attitude by Rousseau and Hayes-Roth) however no provisions are made in the mood or emotional model for the influence of personality. A suite of actions that have personality traits attached to them as conditionals further complicates the model. Rousseau and Hayes-Roth's model does not allow for planning, confining the actor to spur of the moment improvisation (Rousseau and Hayes-Roth, 1998).

In Kshirsagar (2002), the focus of the work is the personification of virtual humans. That is to say, the focus is the endowment of virtual humans with personality. He identifies four types of personification: physical, expressional, logical, and emotional. Physical personification is

defined as designing the way the virtual human looks, while expressional personification is defined as how the virtual human uses gesture and facial expression to express himself. Logical personification is the design of how the virtual human processes input, reasons, and creates output. Emotional personification is defined as designing the "mind" that controls the evolution of emotions over time (Kshirsagar, 2002).

Kshirsagar fulfills emotional personification by implementing a multilayer personality model for synthetic characters in which the Five Factor Model (FFM) of personality, a model of mood, and an emotional model that is loosely based on the Ortony, Clore, and Collins (OCC) model are used. Kshirsagar uses a layered approach similar to Rousseau and Hayes-Roth (1998), claiming that personality is practically immutable over time, that while mood is a prolonged state of mind it can change over time, and that emotions are transitory. Kshirsagar also relates the three layers with personality being the highest, most deliberative layer, mood being affected via personality and emotion and emotional state being instantaneous and affected by mood (Kshirsagar, 2002).

Kshirsagar fulfills logical personification by the alteration of A.L.I.C.E., a classification-based chatbot. A.L.I.C.E is based on a statically-defined pattern-response system that matches natural language statements to syntax patterns stored in memory and mechanically responds (A.L.I.C.E., 2004). This pattern-response system is particularly brittle in the face of uncertainty and responds to these situations by matching more general statements (e.g. if no entry for "I want a piece of cake" exists, but there is an entry for "I want *", the system responds with the defined output for "I want *"). Kshirsagar's modifications to the system include adding a Bayesian weight to the input matching and specifying emotional content to respond with. It is unclear how deeply in the pattern tree these emotional cases were implemented, and it could be particularly

14

problematic if a general statement such as "I want *" has an attached emotional context since "I want a piece of cake" and "I want to laugh until my sides hurt" have completely different emotional connotations. Also, while this model relies on tracking the personality, mood, and emotion of the synthetic character, what is actually expressed is the personality and emotional preferences of the designer (Kshirsagar, 2002).

Each of the methods presented above is incomplete in some respect, and while the methods are foundational, even if the methods were complete, they would not be sufficient to complete the proposed work. Chen et al.'s model offers real time, or slightly delayed, motivation and emotion synthesis which in turn influence behavior, however, it seems to be firmly locked in the present, excluding such emotions as anticipation or dread. Gratch and Marsella adequately address this shortcoming, but fail to adequately address how behavior impacts emotions and does not adequately allow for the agent to learn from past emotional experiences. El Nasir et al.'s method seems well suited to allow the past to influence of emotional state, but does not seem to address the correlation of emotion and intellect. These models only address the inclusion of an artificial emotive state in planning and goal fulfillment. All of the models appear to have been developed in absence of relevant psychological theories. In short, the methods proposed by these studies are only part of the equation.

Rousseau and Hayes-Roth's method allows for the influence of personality, mood, and emotion on the actions of the synthetic character, but does not allow their influence on each other. Kshirsagar's model is the most complete, although the implementation is very limited and does not allow for the expression of the personality and emotional model by the character. Both of these systems are also unsuitable for displaying lexical and creative intentionality as implemented.

<u>Intentionality In Synthetic Characters</u>

Humans find interaction with synthetic characters to be the most synergistic when they feel an affinity with the character's mental state based on the observable behavior of the character. Therefore, for the purposes of the present work, let intentionality in synthetic characters be defined as "the perceived link between a character's actions and its beliefs and desires." However, it is equally engaging to interact with characters for which we feel no affinity with the character's mental state, such as villains in a work of fiction, although we must still glimpse the motivations behind the behaviors. The important factor that must be conveyed via intentionality is the believable emotional expression and other social aspects of the character derived from the situation (Paiva et al., 2001; Rousseau and Hayes-Roth, 1998).

Characters in traditional media fulfill a dramatic role, such as that of a hero, or that of a villain. Roles are defined as a generalized class of individuals for which prototypical behaviors, relationships and interactions are known both to the actors and the audience (Hayes-Roth, 1997). Creating believable, yet synthetic, emotional expression is hinged upon working within the established character role and making role appropriate responses to the user's attempts to change the interaction. In this case, the 'role' is defined by the defined personality, i.e., the emotional and social aspects that lend believability to the performance of the synthetic character (Rousseau and Hayes-Roth, 1998). These dramatic roles help humans to construct a mental model by conveying prototypical behaviors, schema (McKeown, 1985), scripts (Shank and Abelman, 1977), and the like.

In order to construct a mental model of a synthetic character's mental state, one must feel he gains insight into the motivations and thought processes of the character based on visual

communication, linguistic communication, or a combination of the two. In this case, visual communication refers to actions, quality of action, timing of action, form-appropriate gesture, expression, and other visual cues of mental state appropriate to the character (Yoon et al., 2000). Linguistic communication may be textual or auditory, although auditory implementations must also address intonation and tempo.

Traditionally, humans communicate through language, gesture, gaze, expression and touch. Gestures that do not contain qualitative aspects of movement do not convey motivation and personality (Badler et al., 2000). Communicative gestures have been cataloged into five general types:

1. Iconics – representative of some feature of the subject, such as its shape or orientation.

2. Metaphorics – representative of abstract features of the subject, such as its use or exchange.

3. Deictics -- used to indicate some spatial point.

4. Beats -- hand movements that accent spoken words and speaker turn-taking.

5. Emblems -- stereotypical patterns with well understood semantics. (Badler et al., 2000)

As mentioned above, communication between humans may be either lexical or creative in nature (Poggi, 2001). A lexical system is one that relies on a lexicon of syntax-semantics pairs, while a creative system is one that relies on a set of encoded rules to create new syntax for a given meaning or to decipher a meaning from a given syntax. Any lexical system and any creative system developed for a synthetic character must operate within the scope of the established role of the character and must be appropriate within the scope of the situation.

Equipping a synthetic character with lexical verbal skills to implement intentionality is necessary, but clearly not sufficient. A.L.I.C.E. is a text-based chatbot system that operates

17

exclusively from the realm of lexical communication, based on the erroneous belief that human communication is the result of only pattern-response behavioral learning (A.L.I.C.E., 2004). Using A.L.I.C.E. can be entertaining, and engaging due to the personality of the "bot master" entering the pattern-response pairs into the lexicon, but the system breaks down in the face of uncertainty, often replying with nonsensical phrases, broken sentences, repeated sentences, or general responses such as "I didn't get that. Try being more specific or less specific." (A.L.I.C.E., 2004). Lexical-only systems do not fully meet the criteria of adaptability.

Endowing a synthetic character with the means of conveying lexical and creative intentionality allows the synthetic character to reliably express its mental state, even when faced with concepts that are outside its lexicon. While the combination of gesture, gaze, and verbal language is the most human-like method of face-to-face communication, any single form is sufficient to convey intentionality, albeit with less precision and a greater possibility of misunderstandings. If the creative system is expanded to allow the character to transform lexical responses to match its mood, emotional state, or personality factors, and provided that the mood, emotion, and personality models are implemented with strict attention to psychological research, the synthetic character will be able to reliably express its mental state in a believable and consistent manner.

Personality And Personality Dynamics In Synthetic Characters

The author feels strongly that in order to model human personality with enough accuracy that humans feel affinity with the mental state of the synthetic character; one must base the model on an accurate representations and relevant theories of personality. Additionally, one

must give sufficient thought to personality dynamics, that is, "the mechanism by which personality is expressed, often focusing on the motivations that direct behavior" (Cloninger, 2004). Personality dynamics is tightly coupled with the author's definition of intentionality. Individuals may be identified by assessing and quantifying their cognitive traits and, since experimental analysis shows a systemic and complex relationship between non-cognitive traits and cognitive performance (Revelle, 1995), by identifying their personality in accordance with an established personality theory. Personality theories are mainly concerned with the structure of the human mind, or psyche, which subsumes the explanation of how individual psychological processes are organized and made coherent.

The author submits that in order to create a synthetic personality there are seven necessary and sufficient conditions:

1. a method of description of the uniqueness or individual characteristics that set the synthetic personality apart from similar creations

2. a descriptive characterization of the behavior of the construct

3. a predicative characterization of the behavior of the creation

4. the subjective sense of internal consistency or the 'memory of identity' of the synthetic personality

5. the methods of motivation, that is free-will vs. deterministic behavior

6. the method of lasting change to the personality

7. the method for growth

Implied by these conditions is the concept of identification, that is, the ability of others to guess the personality type given to the character. Therefore, the personality dynamics of the

character must also be identifiable, i.e., the emotive state, the mood, and the cognitive state of the character.

Carolus Linneaus said: "All real knowledge that we possess depends on the methods by which we distinguish the similar from the dissimilar" (Fujita, 2004). Thus, the question presents itself: "Which personality theory is best suited for the creation of synthetic personality?" We must not use just any personality theory when creating synthetic personalities, but the type of theory that is best suited for the work. Considerable thought must be given to the questions of design and the intent of the construct, and comparisons should be made between the theory and the cognitive model to be implemented as well.

Over the years, several schools of personality theory have developed, including: psychodynamic theories espoused by Freud, Jung, Adler, Eriksson, and others; humanistic theories like those associated with Rogers, Maslow, and others; behavioral theories such as those identified by Skinner, Staats, and others; and finally, trait or taxonomic theories hypothesized by Allport, Catell, McRae and Costa, and others. These approaches vary widely in almost every detail, and all offer valid and important insights into the human condition.

When the continuum of personality theories is analyzed using the criteria of the scientific method versus humanism, two important approach-based partitions emerge, namely nomothetic-based theories versus idiographic-based theories (Cloninger, 2004). The nomothetic approach studies groups of individuals, and allow us to compare one person with another. The idiographic approach studies individuals without making systematic comparisons to others. Both approaches have strengths and weaknesses and truly understanding human personality will likely require contributions from both approaches (Cloninger, 2004), but for the purposes of the present work,

the nomothetic approach is the best suited as feeling affinity for the mental state of another requires a comparison to one's own mental state as a reference.

Freud's Psychodynamic theory proposes that behavior is the result of psychological forces according to the assumption of psychic determinism. Personality is described in terms of the "id", a structure that unconsciously seeks the immediate fulfillment of biological drives, the "ego", a structure that adapts to reality through the application of defense mechanisms in the face of conflict, and the "superego", a structure that produces guilt and an ego ideal in response to societal restrictions. Psychodynamic theory emphasizes the role of the unconscious and psychosexual development of personality (Cloninger, 2004).

Jung proposes that "the unconsciousness contains broad psychic energy, rather than simply sexual energy" and that consciousness and unconsciousness exist in a relationship of compensation (Cloninger, 2004). In fact, Jung proposes that through the process of "individuation", unconscious aspects of personality are developed and integrated with the consciousness into the mature self. Based on the dimension of introversion and Extroversion, Jungian personality theory espouses eight psychetypes, such as thinking-feeling, and sensing-intuition (Cloninger, 2004).

Adler emphasized conscious striving and goal driven behavior, as well as the creative self. Though Adler thought of each person as unique, his theory listed three types of "mistaken styles of life", such as the getting type, the ruling type, and the avoiding type. He theorized that the cause of these unhealthy styles rested firmly with the style of parenting, i.e., pampering or neglectful. Emphasis is placed on being socially useful and social interest (Cloninger, 2004).

Eriksson proposed an eight-stage theory of psychosocial development. These stages involve culturally relevant crisis resolution for each individual. Culture is the major influence on

21

development and conversely, individuals influence culture through the way they develop at each stage. Eriksson's theory is well supported by research and has been shown to be cross-cultural (Cloninger, 2004)

With respect to the present work, the emphasis on psychosexual development and unconscious processes make Freud's theory unsuitable. Jungian personality theory is also rejected, due to its reliance on individuation, and the emphasis on unconscious processes. Adler's theory relies too much on parental influence to be a viable model for a synthetic character. Eriksson's theory is also unsuitable since it relies on a mutually influential relationship between the individual and his culture. All of these theories are highly deterministic. Therefore, psychodynamic theories of personality will not be used in the present work.

Rogers proposed a theory of personality based on the humanistic viewpoint in which the individual actively seeks higher development, motivated by the "actualizing tendency" as opposed to development via passive determination by external or unconscious forces. Rogerian theory was specified so that it could be empirically verified, but focused more on personality change than on the development or structure of personality. Rogers did not offer a scheme for understanding personality differences (Cloninger, 2004).

Maslow proposed a theory of personality that was less deterministic and more focused on human values. Maslow's Need Hierarchy identifies five levels of development ranging from physical to emotional to psychological. Individuals are motivated by deficiencies through the first four levels and once they become "self-actualized" the source of motivation switches to "being motivation", i.e., the need to fulfill their potential. The theory emphasizes immediate experience as an aspect of psychological health (Cloninger, 2004)

22

The lack of a scheme for differentiating personality makes Roger's theory unsuitable for the present work. The subjective nature of Maslow's theory causes it to be abandoned. Also, given that humanistic theories are idiographic approaches (Cloninger, 2004), they will not be suitable models for the current work.

B.F. Skinner proposed a theory of individual behavior in terms of only observable, environmental determinants. Internal characteristics or traits were ignored. Less a personality theory, than an alternate theory to explain human behavior, Skinner's theory of "operant conditioning" proposes that behaviors are acquired as a result of reinforcement and extinguished as a result of punishment (Cloninger, 2004).

Staats proposed a theory of psychological behaviorism that attempts to address some of the inadequacies of Skinner's work. He proposes that personality is composed of three categories of "basic behavioral repertoires" (learned behaviors that have broad effects on personality and are the basis for later learning), namely: the emotional-motivational repertoire, the language-cognitive repertoire, and the sensory-motor repertoire. Psychological behaviorism describes personality as learned, in which effective learning produces psychological adjustment and ineffective learning produces maladjustment (Cloninger, 2004).

Skinner's theory is criticized as being overly deterministic and reductionist (Cloninger, 2004) and fails to address the issue of human individuality, which is at the center of the proposed work, and is thus unsuitable. Conversely, Staats' psychological behaviorism theory seems to fit well with the concept of lexical intentionality, as it seems to describe a non-deterministic lexicon of pattern-response pairs that are conveniently similar to the needs of a synthetic character and that can be learned and extinguished. Unfortunately, although it does embrace the idea of identification of individual through psychological testing, Staats' work alone does not seem to be

sufficient to adequately differentiate an individual, but will be used in the present work to help fulfill lexical intentionality with regard to personality dynamics.

Allport defines personality as "the dynamic organization within the individual of those psychophysical systems that determine his unique adjustments to the environment" (Cloninger, 2004) and identifies the atomic unit of personality as the trait. Evidence to support his claim of personality traits comes from many sources, including language, behavior, studies, etc. Allport describes three kinds of traits: "Cardinal traits" that have pervasive influence but rare occurrence, "Central traits" that have broad influence and ubiquitous occurrence, and "Secondary traits" that influence only a few behaviors (Cloninger, 2004).

Cattell defined personality as "that which permits a prediction of what a person will do in a given situation" and developed many different types of personality tests and factor analyses seeking the underlying source traits that determine personality. He distinguished several types of traits including: dynamic, temperamental, and ability, and arrived at five fundamental traits through a second-order factor analysis of the results of his personality testing (Cloninger, 2004).

With regard to synthetic characters, any of the trait-based taxonomic theories fit well. First, consider that the purpose of a taxonomy is to describe the universe of phenomena to be explained or studied (Fujita, 2004). Fujita (2004) goes on to say: "A good taxonomy will point researchers in fruitful directions in their research and will allow a standard vocabulary by which their research results may be communicated and related with each other." Although synthetic personalities are proposed by the author primarily as a method of increasing user immersion and engagement, psychologists should be able to evaluate a synthetic personality based on relevant personality theory and find some parallels to human behavior. Additionally, a character that is

built from sound psychological principles could be more effectively used in psychological research and simulations.

There are several personality taxonomies to choose from, including Allport's theory, and Cattell's five trait theory, but due to the amount of existing research on McRae and Costa's Five Factor Model (FFM), it is the most likely candidate. The structure of the FFM itself also helps to distinguish the taxonomy as the appropriate choice, given the high degree of coupling between the model and agent-based reasoning, a common architecture for implementing synthetic characters.

The FFM breaks personality into five continua:

1. Neuroticism

2. Extroversion

3. Openness

4. Agreeableness

5. Conscientiousness (Costa and Widiger, 2002)

These continua help to identify personality by locating the individual in a normal, an abnormally high or an abnormally low range (Clark et al., 2002). This feature directly relates to numbers one, two, three, and four of the seven necessary and sufficient conditions for synthetic personality as defined by the author above. Specifically, a large amount of psychological research has shown that personality classification based on the FFM can be predictive in terms of responses from individuals (Costa and Widiger, 2002). Also, the very fact that the FFM is used to classify individuals as a specific personality type helps to mark one individual from the next. Finally, the FFM is being used as a characterization of behavior in that it is gaining acceptance as a diagnostic tool for the Diagnostic and Statistical Manual (DSM) (Clark et al., 2002), which

while not directly relevant to the planned work, only emphasizes the merit of using synthetic personalities in psychological research. Further, according to Costa and Widiger (2002), distinguishing normal and abnormal psychology makes no sense from a trait-based approach, and therefore the diagnoses listed in the DSM may be used as additional identifiers for individual characters.

Neuroticism is defined as "a chronic level of emotional adjustment and instability." High measures of Neuroticism identify subjects who are prone to psychological distress. These individuals commonly display a "negative affect, unrealistic ideas, excessive cravings, difficulty tolerating frustration caused by not acting on their cravings, and a maladaptive coping response" (Costa and Widiger, 2002).

Extroversion is "representative of the quantity and intensity of preferred interpersonal interactions, activity level, need for stimulation, and capacity for joy." A person measuring high on the Extroversion scale is likely to be "sociable, active, talkative, person-oriented, fun-loving, and optimistic". Conversely, a person measuring low on the Extroversion scale is likely to be "reserved, sober, aloof, independent, and quiet" (Costa and Widiger, 2002).

Openness involves "active seeking and appreciation of experiences for their own sake." Open individuals are "curious, imaginative, and willing to entertain novel ideas and unconventional values." They experience the whole gamut of emotions more vividly than do closed (those with low Openness measures) individuals. Closed individuals are often "conservative, conventional, dogmatic, rigid in their beliefs, set in their ways, and emotionally unresponsive" (Costa and Widiger, 2002).

Agreeableness is "an interpersonal dimension that refers to the kinds of interactions a person prefers along the continuum from compassion to antagonism." Highly agreeable

individuals are often "soft-hearted, good natured, trusting, helpful, forgiving, and altruistic." They tend to be responsive and empathic, and believe most others are cut from the same cloth. Individuals with a low Agreeableness measure are likely to be "cynical, rude, abrasive, suspicious, uncooperative, irritable, and can be manipulative, vengeful, and ruthless" (Costa and Widiger, 2002).

The Conscientiousness domain "assesses the degree of organization, persistence, control, and motivation in goal directed behavior." People with a high measure of Conscientiousness tend to be "organized, reliable, hard working, self-directed, punctual, scrupulous, ambitious, and persevering." Low measures of Conscientiousness often indicate "aimless, unreliable, lazy, careless, lax, negligent, and hedonistic" individuals (Costa and Widiger, 2002).

Taxonomic theorists also address the question of free will versus determinism. Traits are viewed as individually consistent adaptations to the environment rather than fixed-point behavioral responses (Clark et al., 2002). In other words, each response that is based on or influenced by a personality trait is dynamic in nature and can be altered directly by the environment or the previous actions of the individual. This meets conditions five, six, and seven as defined above.

The FFM is as good a taxonomy as can be expected until a more coherent physiology of personality processes is achieved (Fujita, 2004). Further, the FFM is lexical in nature, that is, it assumes that socially relevant personality characteristics have become encoded in our language (Fujita, 2004). Also, and perhaps most importantly, if the FFM describes basic personality traits, they should be trans-cultural (Fujita, 2004). Finally, it is a hierarchical model of the structure of personality traits (Costa and Widiger, 2002).

In Kshirsagar (2002), a multilayer personality model is proposed for synthetic characters in which the Five Factor Personality model, a model of mood, and an emotional model based on Ekman's six basic emotions are used. Kshirsagar uses a layered approach, claiming that personality is practically immutable over time, mood can change over time, but is a prolonged state of mind, and emotions are transitory. Kshirsagar also relates the three layers with personality being the highest, most deliberative layer, mood being effectible via both personality and emotion, and emotional state being instantaneous and affected by mood (Kshirsagar, 2002).

Kshirsagar's model seems inadequate and overly simplified. Vel'asquez (1997) differentiates mood and emotion only by level of arousal, rather than length of activation. Additionally, Gross (2002) differentiates affect across four valence states: stress response, emotion (such as anger and sadness), emotion episodes (such as tirades and bar-room brawls), and moods (such as depression and euphoria). That the FFM of personality is tightly coupled with emotion can hardly be argued, given that the Neuroticism domain is the measure of emotional stability versus emotional lability, and at least two of the other domains, Extroversion, Openness, have emotional blandness as a possible outcome of low measures. Finally, with respect to the FFM itself, personality theorists view traits not as fixed-point behavioral responses, but as reflecting adaptations to the environment that are consistent with each individual (Clark et al., 2002) and are therefore not immutable, but dynamic.

Kshirsagar implements a Bayesian Belief Network (BBN) for each factor of the FFM, however these BBNs are used only to calculate the present mood of the character given its latest mood and the normalized values from the personality model (Kshirsagar, 2002). This, however, seems to clash with the mood research presented above, the emotional models examined, and the FFM itself. Further, Kshirsagar does not allow for any adaptation, believable or otherwise, of

the response generated, nor does it allow for the believable expression of the character's state of mind, but relies on pre-scripted "emotional responses" that are written by the developer and associated with some emotional state (Kshirsagar, 2002).

Kshirsagar implements his personality model without regard for the proposed necessary and sufficient conditions of synthetic personality. Items two, three, six, and seven are not fulfilled by Kshirsagar's model. Although the model should provide a descriptive characterization of the behavior of the character, Kshirsagar's implementation limits this characterization to a change in emotional state, since the behavior, as represented by the response from the character, is pre-scripted by the designer and is therefore descriptive only of the designer. The same argument applies to predictive characterization. Kshirsagar outlines a partial method that might be used to allow for growth and lasting personality change, but the mere exercise in predicting the next mood or emotion is useless unless some evaluative measure is used to determine the success of the prediction. This evaluative measure could be extended to allow for growth and lasting personality change if the measure were applied to the weights in the BBN.

Rousseau and Hayes-Roth's (1998) "social-psychology model" of personality is based on trait-based personality theory and social learning theories. Rousseau and Hayes-Roth fail to identify which trait-based theory is used, stating only: "Unfortunately, there are many theories proposing various personality traits and none of them are universally accepted" (Rousseau and Hayes-Roth's, 1998). Indeed, it is important to note that as Cloninger (2004) points out, "there is no single paradigm to serve as a theoretical model accepted by the entire field of personality." However, Revelle (1995) states that there is a consensus "among most, but not all description taxonomists" that the FFM is the best trait-based theory. As mentioned earlier, Fujita (2004)

states that the FFM is as good a taxonomy as can be expected until more coherent physiology of personality processes can be achieved. Further, many researchers have settled on five basic personality traits through various lexical criteria (Fujita, 2004; Cloninger, 2004; Kokkonen and Pulkkinen, 2001).

Rousseau and Hayes-Roth's (1998) model implements traits on a numeric scale and the traits are subject to very little change over time. They implement mood and emotion on a similar scale, but allow for more change at a more rapid pace. Rousseau and Hayes-Roth's (1998) model is implemented to support all seven of the necessary and sufficient conditions for synthetic personality as described above.

Personality in synthetic characters is pointless without the believable expression of emotions, mood, and individual characteristics, i.e., personality dynamics. As can be seen in the review of the synthetic personality implementations above, the definitions of emotion and mood vary widely from implementation to implementation. This may be the result of the vacillation of psychologists on the subject and the fact that the term emotion is a term that was lifted from common language with all of its connotations intact (Gross, 2002). Recently, researchers have returned to the concept of emotion and have postulated prototype conceptions of emotions (Gross, 2002).

Based on the work of Ekman, Russell, and others, three noteworthy features of prototypical emotion exist:

1. Emotions arise when an individual evaluates a situation as significant.
2. Emotions are multifaceted, whole body processes that involve changes in the domains of subjective experience, behavioral expression, and central and peripheral physiology.

30

3. Emotions may be described as either dimensions or categories (Gross, 2002)

By implication of the third feature, definition of mood may be unnecessary. Indeed, if one adopts the dimensional approach to emotion, relatively broad dimensions that describe changes in the emotional experience, expression, and physiology are adopted. Emotions are seen as continuously distributed over a few dimensions, such as positive affect, negative affect, intensity, pleasantness, etc. This seems to fit with the work of Vel'asquez (1997) and with Buddhist philosophy (Solomon, 2002).

Conversely, the categorical approach focuses on discrete emotions and differentiates between emotions. Descartes proposed that six simple and primitive passions exists: wonder, love, hatred, desire, joy, and sadness and went on to say that all other emotions are compositions of some number of these six or subspecies of them (Solomon, 2002). Other philosophers, such as Aristotle, Hobbes, Spinoza, Hume, Rousseau, and Smith also espoused some number of "basic passions" (Solomon, 2002). Ekman (1992) identified six basic emotions (also called affect programs): joy, anger, fear, sadness, disgust, and surprise. Despite the apparent popularity and historical precedence of this categorical approach, this idea remains controversial; the dispute is centered on the concept that a small set of basic emotions exist (Hudlicka, 2002). Indeed, the point of Solomon's (2002) paper is to caution against such "cognitive as well as biological reductionism", claiming that emotion is a "holistic" experience and that any emphasis on one aspect or another distorts the subject.

Other theories of emotion exist, such as the social construction view of emotion as championed by Averill, which argues that emotion is characteristically the result of social interaction, Oatley's computational model, and Ortony, Clore, and Collin's cognitive model of

emotion (Watts, 1997). The author has avoided a review of "mind-body" theories due to the focus of the proposed work.

Research examining the link between personality traits and emotions is of particular pertinence to the proposed work. Plutchik goes so far as to propose that traits are the result of a combination of emotions and that humans identify the personality of others based on emotional behavior (Emmons and Deiner, 1986). Similarly, it has been found that traits and emotions can be represented in the similar circular arrangements based on the dimensional approach, although the dimensions used in the personality circumplex vary from theorist to theorist, while the emotion circumplex is generally constructed around the Hedonic Tone and Activation axes (O'Malley and Gillet, 1984).

O'Malley and Gillet's (1984) work goes on to show that there is an identifiable, systematic relation between personality traits and emotions. Indeed, O'Malley and Gillet state: "...emotions are fairly well determinable from specific traits." Though O'Malley and Gillet did not correlate all five traits of the FFM, their research used Neuroticism-Stable and Extroverted-Introverted traits and developed roughly the same list of emotions as specified in the discussion of the FFM above.

Emmons and Deiner (1986) performed a similar study in the context of "ecologically valid settings, i.e., individuals everyday lives" and report theoretically meaningful relations between various personality dimensions and specific emotions. The study also supports the inclusion of Staats' psychological behaviorism theory with a trait-based model of psychology in that the researchers found that the experience of certain emotions are not always independent of context (Emmons and Deiner, 1986).

As this short review indicates, psychological, philosophical, and social research in emotion is particularly chaotic, with little in the way of consensus or prevailing theories supported by empirical research. Gross (2002) speculates that this is the result of emotion researchers emphasizing differences among approaches and definitions. One concept that each of these theories share is that emotions have an antecedent and a response. Many theories emphasize that emotion is governed by an appraisal of the antecedent and that the process of emotion regulation influences emotion (Gross, 2002). Gross (2002) goes on to state that emotion regulation is necessary due to the impact of our emotions on a wide range of mental processes. Emotion regulation includes the concepts of expression of emotion and the subjective experience of emotion (Kokkonen and Pulkkinen, 2001).

As with emotion, precise definition of mood is elusive. Two approaches are generally taken in regard to mood and emotion. The first approach is most often used and views mood and emotion as conceptually equivalent, and therefore interchangeable. The second approach, which was used in both of the synthetic characters that implemented a personality and emotion model reviewed above, differentiates mood and emotion mostly by specification and duration. Emotions are generally considered to be brief, event driven, object related, and relatively intense. Emotions are also associated with distinct expressions and influence action. Mood is generally considered broader and longer acting, and of lower intensity. Mood is also assumed to influence cognition rather than action (Kokkonen and Pulkkinen, 2001).

The author finds both of these definitions insufficient. First, if we assume that mood and emotion are conceptually equal, they should not be distinguishable from one another. If that is true, then it must not be possible to be in negative mood and feel happiness, or in a positive mood and feel any negative axis emotion, but experience with the human condition rapidly

33

disproves this argument in the author's opinion. Therefore, the first approach to differentiating

mood and emotion must be flawed. Second, if the second method is correct, there can't be a

single example that fits both categories. By counter-example, grief can be catastrophic in terms

of intensity and duration. It can influence action, and cognition. It can also be associated with

distinct expressions, and is event driven. It may be object related (such as the grief felt at the

loss of a loved one) or may not be object related (for example, the grief derived from failing an

important life goal, such as completing a dissertation). Grief must be both an emotion and a

mood, and therefore, the second definition listed above is also flawed.

For the purposes of the present work, let mood be defined as meta-emotional experiences

and let emotion be defined as the class of cognates (atomic units of cognition) that correspond to

distinct feelings such as joy, anger, etc. Further, the author proposes that mood acts as a partition

on the class of emotions, in other words, segregates emotions into like groups. For instance,

positive and negative affect segregate the class of human emotions into two distinct groups, with

emotions like joy, love, pleasure, delight, and serenity in one equivalence class and emotions like

anger, hatred, disgust, and lethargy in the other. Thus, it can be argued that emotional

experience is classified hierarchically, a position supported by the theories and research of

Deiner, Smith, and Fujita; Russell and Feldman Barrett; and Watson and Clark (Gross, 2002).

Both mood and emotion must have discrete state models in a synthetic character and both

must be able to influence the other equally. For example, the author can identify a positive mood

and still be experiencing a negative axis emotion. If the negative axis emotion continues to be

reinforced, the author's mood may gradually become negative, while a negative emotion that is

not reinforced may fade in the face of the positive mood. It certainly can be argued that

numerous reasons for this exist and the author makes no claims that this anecdotal example is proof of any theory.

In summary, to implement personality and personality dynamics in synthetic characters, it is necessary to identify the appropriate model of personality, identify a suitable theory of emotion, identify a suitable theory of mood, and allow all three of these models to influence one another equally. Further, the synthetic character's personality must be identifiable based on the expression of its emotion and mood. Additionally, these three models must integrate the model of cognition used in the character.

Cognition In Synthetic Characters

What is cognition? This is a difficult question to answer since the processes under study are internal and are often studied using retrospective reports which may not be trustworthy (Hurlburt et al., 2002). Newell and Simon argue that the human capacity of thought, emotion, communication, and behavior arise from our capacity to process symbols (Green et al., 1996). Searle challenged the idea with his "Chinese room argument", and attempted to illustrate that thinking can't be limited to symbol manipulation alone (Green et al., 1996). Taking this argument a step farther, Harnad proposes that certain basic symbols are linked to representations of objects, and thus cognition is best modeled using a symbolic-connectionist approach (Green et al., 1996). Harnad's method breaks down in the face of certain linguistic concepts that have no real-world objects, such as permission or obligation.

Whether Newell and Simon are correct or Searle is correct, is beyond the scope of this dissertation. The author is concerned with the believable expression of mental state, and the

question "Is modeling cognition necessary to implement lexical and creative intentionality?" presents itself. This question can be restated as: "Is there a link between cognition and behavior?"

The immediate, intuitive answer is in the affirmative. The link between cognition and behavior is one of the tenets of cognitive psychology (Ashcraft, 1998; Hurlburt et al., 2002), and this link was verified empirically by Hurlburt et al. (2002). Therefore, to implement intentionality, synthetic characters must implement some model of cognition, as well as a model of personality, emotion, and mood.

Previously, cognition was defined as "the collection of mental processes and activities used in perceiving, learning, remembering, thinking, and understanding, and the act of using those processes" (Ashcraft, 1998). To put it another way, cognition refers to all of the human mental processes, "from perception and recognition up through reasoning and inference" (Ashcraft, 1998). In synthetic characters, only a subset of these processes may be necessary.

Implementing lexical and creative intentionality in synthetic characters requires a subset of cognitive processes that is unique from synthetic characters that only implement lexical intentionality or no intentionality. Loosely based on Palinscar and Brown's model of reading comprehension (Hillerbrand, 1989), the author submits that creative intent implies seven explicit cognitive abilities, namely:

1. recognition of and appropriate response generation for user input in an imperfect and incomplete problem domain

2. such processing of user input as required for internal consistency and other cognitive processes outlined here

3. activation of knowledge in memory

36

4. domain-appropriate reasoning

5. evaluation of cognitive and emotion operations, and generative responses for accuracy and appropriateness

6. planning, adaptation, anticipation

7. encoding new knowledge into memory.

Recognition of user input and generation of an appropriate response may sometimes be fulfilled with no reasoning required. Similar to the computer program alluded to in Searle's Chinese room argument, the A.L.I.C.E. chat bot relies on the pattern-response process to mimic conversation, and therefore, no thinking or understanding occurs (A.L.I.C.E., 2004). While many researchers dispute the validity of this process, Ashcraft (1998) states that some human behaviors support this approach, at least those linked with highly over-learned and practiced cognitive processes. These types of processes are said to be "automatic." It can be argued that the antecedents of these cognitive processes have been stored with the appropriate response in a lexicon similar to the basic behavioral repertoires proposed by Staats (Cloninger, 2004). These automatic processes no longer appear to involve actual processing beyond an appraisal of the antecedent and a retrieval of the response, or the processing occurs at such a blindingly fast speed that it is appropriate to model them in an appraisal-retrieval system.

For example, when solving simple multiplication problems (i.e., 2 x 3), it takes the average college student 700-thousandths of a second to respond with the correct answer, while other single digit problems that are perceived as harder (i.e., 6 x 9) take up to a second (Ashcraft, 1998). Ashcraft (1998) estimates that the process of encoding a simple stimulus and responding to it takes approximately 250-thousandths of a second (a duration that is very close to the base response time), which leaves very little time to perform the multiplication itself.

Ashcraft (1998) argues that the retrieval of solutions to these over-learned and practiced problems occurs without conscience thought, and may be simply long-term memory access, or performing the calculation in some optimized neural circuit.

A lexicon of automatic cognitive processes does not paint a complete picture of cognition, however, nor does it allow a synthetic character to recognize and respond to all forms of input in an incomplete problem domain. There are many cognitive processes that are not automatic, in fact they are quite slow and deliberate (Ashcraft, 1998). Although certain inferences (such as the answer to the question "Does fire burn?") can become automatic, many others (such as the answer to the question "Did Zeno choose to live by the Four Cardinal Virtues: Wisdom, Justice, Courage, and Decorum?") do not become automatic due to their relative rarity or complexity. Therefore, in addition to an appraisal-response model, synthetic characters implementing creative intentionality must also model these deliberate, conscious cognitions that occur in the face of unfamiliar problems.

Though there are many different theories about human problem solving, including Gestalt theory (Ashcraft, 1998), solving problems by insight (Ashcraft, 1998), analogical problem solving (Green et al., 1996; Ashcraft, 1998), the multi-constraint theory (Ashcraft, 1998), and the expert model (Green et al.,1998), Ashcraft (1998) identifies four general characteristics of problem solving:

1. goal directedness
2. a step-wise solution
3. cognitive operations
4. sub-goal decomposition.

The Gestalt theory relies on two basic principles: (a) humans perceive "wholes" or patterns rather than individual parts, and (b) the whole is different from, or greater than the sum of its individual parts. In his work with animals, Kohler generalized these principles to say that relations among stimuli are perceived rather than individual elements of the stimuli. Kohler called this sudden perception of relations "insight." Additionally, two specific difficulties in Gestalt problem solving were made by the study of humans: the tendency to use objects in ordinary rather than creative ways during problem solving, called functional fixedness, and the tendency to try to solve new problems with old strategies rather than adapting them to the new situation called negative set, or set effects (Ashcraft, 1998).

Insight can be defined as "a deep, useful understanding of the nature of something" (Ashcraft, 1998) and often occurs suddenly. Sternberg attributes the sudden nature of insight to the novel interpretation of the problem or the application of a novel approach. It may be just as valid, however, to think of insight in terms of overcoming functional fixedness, negative set, or examining the problem from a different perspective. Wickeldon states that insight sometimes means that a critical inference that leads to the solution has been drawn (Ashcraft, 1998).

Analogy is the "relationship between two similar situations, problems, or concepts" (Ashcraft, 1998). Often, analogical problem solving is used to solve problems for which there is inadequate knowledge (Green et al., 1996). For example, one might think of network connections as pipes of water, or electricity moving through wires as crowds of people moving through passageways. For an analogy to be useful, the two problems must be put into some kind of relationship so that the similarities or differences can be seen (Ashcraft, 1998). Gick and Holyoak performed analogical problem solving research on humans and discovered that 20% of their subjects were able to spontaneously notice and use an analogy for their problem, but once

39

they were prompted to use it 92% of their subjects were quite good at mapping the analogy to the new problem (Green et al. 1996; Ashcraft, 1998). Green et al. (1996) notes that this difference between noticing analogies and mapping known analogies suggests that individuals may not be able to solve certain problems due to the inability to access suitable analogies.

The multi-constraint theory is an overall theory of analogical problem solving based on such research as Gick and Holyoak's. Proposed by Holyoak and Thagard, the theory predicts how humans use analogies in problem solving and what factors govern the analogies that humans construct. The theory states that the use or development of analogies is constrained by three factors: problem similarity (i.e., a reasonable degree of similarity between the source domain and the target domain), problem structure (i.e., the one-to-one mappings between the source and target domains), and the purpose of the analogy (Ashcraft, 1998).

The expertise model, proposed by Newell and Simon, describes problem solving as a state space search through a problem domain. In this view, a mental model is a state within the problem domain and operations on the mental model are equivalent to traversing the space. Problems are viewed as a goal and a set of sub-goals. The process of transforming an initial state into a goal state can be viewed as a set of production rules, or productions. Productions are simply logical implications, most commonly expressed as IF <predicate> THEN <response>, where the <predicate> is the local goal and any preconditions and the response is the operation that transforms the state (Green et al., 1996).

Humans who are experts in a particular problem domain are able to reduce computational complexity of problems in that domain by manipulating incoming information into recognizable patterns. Expert knowledge is represented by organized patterns or "schemas" in memory. These schemas contain knowledge about specific problems, situations, and relevant knowledge

about the expert's skills that could be applied to the problem. Experts are able to make broad inferences, unify seemingly disparate problems by recognizing underlying features, and make "qualitatively more sophisticated, critical judgments." Experts also have metacognitive abilities, that is, cognitive abilities that are applied to the expert's cognitions. Metacognitive abilities include the ability to evaluate reasoning, plan, monitor processes, etc. (Hillerbrand, 1989)

Gestalt theory is impractical to implement in synthetic characters as it seems to rely on high level cognitive processes, such as the interpretation of patterns and the identification of complex relationships. The insight model could be used in a synthetic character, provided that Wickeldon's definition of insight is used. This definition, however, seems to restrict insight to a natural outcome of the model of expert problem solving. The expert model can certainly be implemented, indeed, Gonzalez (2002) claims that the problem of creating a computational expert system is a solved one. Expertise is a possible solution to the deliberate cognition problem.   The analogical problem solving method and constraint theory can both be implemented in synthetic characters and either method also presents a possible solution to the problem.

Newell and Simon proposed the first computational model for problem, named the General Problem Solver (GPS), that was designed to implement the most significant problem solving heuristic called "means-ends analysis" (Ashcraft, 1998) and that relies on production rules to transform one problem state into another. In means-ends analysis, the current state is compared to the stated goal to detect differences between the two. Once the differences are identified, an operator that reduces the difference is searched for. Failure to find an operator capable of being applied to the current state initiates a process of setting up a sub-problem that would allow transition to a state that the operator could be applied to. Means-ends analysis can

then be applied recursively to reduce all differences that can be identified, thus allowing the goal state to be reached. GPS was not limited to solving problems in a specific domain; it was able to solve any problems for which means-ends analysis was appropriate. While it was a very good simulation of human problem solving, there were cases in which the modeling was low fidelity, specifically for problems in which the application of means-ends analysis was questionable (Ashcraft, 1998; Luger and Stubblefield, 1993; Rich and Knight, 1991).

Anderson proposed another general model called ACT* (adaptive control of thought) which covers the entire array of cognitive processes, but can also be used as a problem solving model. ACT* incorporates three major components: declarative memory, production memory, and working memory. Declarative memory is a memory network that is accessed by means of "spreading activation". Spreading activation is the process of retrieving concepts based on activation of the concept by reference, and the spreading of this activation to like nodes in the semantic network. The strength of the activation weakens with distance from the original concept and decays over time. Production memory is the stored collection of IF-THEN rules that run the system. The process of matching input to the conditional part of the production rule occurs in working memory. Working memory is the part of the system that "keeps track" of the current state and makes responses to the outside world. ACT* assumes a means-ends approach to problem solving and is supported by empirical testing in human subjects (Ashcraft, 1998).

SOAR attempts to model human cognition by extending the production rule architecture (Luger and Stubblefield, 1993) and is based on a set of specific, cognitively motivated hypotheses about the nature of human problem solving (Rich and Knight, 1991). In SOAR, problems are solved by defining and searching a problem space (Luger and Stubblefield, 1993). In the SOAR architecture, long-term memory is stored as a set of productions, and short-term

memory is a perception affected buffer that stores facts deduced by rules in long term memory (Rich and Knight, 1991). All problem solving in the model occurs as a state space traversal and all results are remembered for future reference (Rich and Knight, 1991). It is important to note that SOAR treats conflict resolution as a problem in its own right and sets up two problem spaces for each problem: a problem space that maps to the problem at hand, and a problem space design to solve the problem of choosing which productions to run from the first space (Rich and Knight, 1991).

Cognition functions in the domain of mental representation. Thus, in order to produce a synthetic character that appears to cogitate, a model of cognition must be adopted that not only includes the necessary appraisal-retrieval and creative problem-solving systems, but also contains a method of constructing, updating, and accessing a mental representation of the character's world space and the ability to reason about the world. The essential feature of these models is that they symbolize a discrete state of affairs, in other words, a description of the world from the perspective of the character at any given time. Indeed, Craik proposed that the central function of human thought was to create such models of reality (Green et al., 1996).

How accurate do these mental representations need to be? Imagine a world in which a complete understanding of every component in a television, the electrical connections to the power plant, and the method of power generation were necessary to turn on the television. Fortunately, for laymen such depth of knowledge is unnecessary. In fact, human mental models are often merely approximations of the world, or are incomplete or incorrect, but can be highly detailed when necessary (Green et al., 1996). Green et al. (1996) goes on to say that our models are only as specific as they need to be for the purposes at hand. Therefore, the world model of a synthetic character is not required to be complex, complete, or even correct. It merely has to

exist in sufficient detail to allow the character to make decisions, to predict possible outcomes of its behavior, and to plan.

Decision making in synthetic characters must result in believable actions, occur in real-time, and function in the face of an incomplete problem domain (Kille and Warwick, 2001; Yoon et al., 2000). Klein (1997) proposed a model of naturalistic decision making (NDM), the recognition primed decision (RPD) model, that attempts to explain how humans make decisions "under conditions of time pressure, ambiguous information, poorly defined goals, and changing conditions." The RPD model does not address the influence of teams or organizational constraints, but those issues do not apply to the proposed work. The model, indeed, NDM itself, shows a high degree of coupling with the concepts of automatic and deliberate cognitions, outlining three levels of action selection, Simple Match, Diagnose the Situation, and Evaluate a Course of Action (Klein, 1997).

In the RPD Simple Match procedure, when the presenting situation is perceived as typical, or recognized as an analog, this recognition produces four by-products: expectancies, relevant clues, plausible goals, and typical actions. Finally, a course of action is implemented (Klein, 1997).

In the RPD Diagnose the Situation method, the presenting situation is examined to determine if it is prototypical, and if it is recognized as such, the same four by-products are produced and a course of action is implemented. However, if the situation is not recognized as prototypical or there are anomalies noted during matching with the prototype, diagnosis occurs (Klein, 1997).

In the Evaluate a Course of Action method of RPD, when the presenting situation is perceived as typical, or recognized as an analog, this recognition produces four by-products:

expectancies, relevant clues, plausible goals, and some number of actions. The actions are evaluated using mental simulation based on the criteria: "Will it work?" If the evaluation is successful, the action is selected. If the action will work with modification, it is modified and re-evaluated. If the action is unsuitable, the next action is evaluated, or if no actions are left the situation is examined and a new prototype is searched for (Klein, 1997).

Reasoning, problem solving, decision making, and knowledge representation may be accomplished using a system of formal logic (Rich and Knight, 1991; Luger and Stubblefield, 1993), although the application of a heuristic and formal logic allows for a better model of human cognition (Rich and Knight, 1991). Axioms of elementary logic define relationships between entities. Logical implication is one such axiom, in which the truth value associated with some antecedent implies a truth value to some consequence (Luger and Stubblefield, 1993). Formal logic provides a method for drawing certain conclusions in a formal, rigorous manner through rules of inference. A well-formed, or sound, rule of inference is one that always draws conclusions that are valid given the system of logic being used (Luger and Stubblefield, 1993). Mathematical deduction is a sound rule of inference that given $P \rightarrow Q$, and P, allows the conclusion Q to be drawn (Rich and Knight, 1991). Given this definition, it follows that deduction is most likely accomplished using forward chaining, that is processing productions from an initial state to produce valid inferences. Induction is a paradigm for learning that uses observations of a specific instance of a phenomenon to build general rules that explain it (Luger and Stubblefield, 1993). For example, after observing that a black cat crossed one's path immediately before tripping, one might induce that every time one sees are black cat, a trip is eminent. It is an unsound rule of inference (Luger and Stubblefield, 1993) and as in the previous example, the general rules developed may be incorrect. Abduction is another rule of inference,

45

that given: P → Q, and Q, allows the assumption that P is true to be drawn (Luger and Stubblefield, 1993). Given this definition, the intuitive method for producing relevant inferences by abduction is to use backward chaining, that is starting from a goal state and processing production rules in reverse order to determine the necessary initial conditions. Being an unsound rule of inference, assumptions drawn using abduction may or may not be true (Luger and Stubblefield, 1993). Attaching some quantification of the risk of incorrectness to assumptions made with abduction enables abduction to be particularly useful in many incomplete problem domains. A non-monotonic reasoning system, that is, a system in which truth values may change, such as a system designed to facilitate non-scripted interaction with humans, is a natural application area for abduction.

For the purposes of the present work, let prediction be defined as connecting a particular precursor state with its successor state one or more times and the subsequent identification of the probable successor state when initial conditions are similar enough. In synthetic characters, prediction can be used to adapt responses to more appropriate ones, to maintain internal consistency, and by applying the results of the prediction to its mental representation to evaluate its reasoning.

Planning is "the process of analyzing an entire sequence of steps in advance to discover where it will lead before the first step is taken" (Rich and Knight, 1991). In a complete problem domain, that is a domain in which all necessary knowledge is accessible at the time planning occurs, planning may be easily accomplished by brute force using deduction and forward chaining (Rich and Knight, 1991). In incomplete problem domains, however, knowledge of the effects of actions on the environment may not be available, making planning impossible by deduction. In these incomplete domains, it is necessary to perform planning based on

speculative or anticipatory computation. This type of predictive planning uses abduction to plan responses beyond these local areas of incompleteness.

Chen et al. (2001 b) proposes a novel cognitive model based on a logical theory of actions that is able to reason about the character's knowledge, actions, and events. The system allows for actions with duration to be represented, uses prediction and a unique specification language to tailor the actions and events specified in a narrative description in an additive fashion, and a hierarchical planning system (Chen et al., 2001 b).

Chen et al.'s planner uses abductive reasoning via a resolution theorem prover, which offers a high degree of similarity between the specification phase and the implementation phase. In this system, compound actions are equivalent to plans and vice versa. During the execution phase, compound actions are decomposed into atomic behaviors for execution. The compound action is therefore described by a conjunction of atomic behaviors, some of which have not yet been executed. This construction allows the synthetic character to be highly reactive to new stimulus and to verify the predictive nature of the plan at each step (Chen et al., 2001 b).

Chen et al.'s (2001 b) system interweaves perceiving the environment, planning, and acting on the environment in a single, common process. Sensor events serve as preconditions for successor states, choice control variables, or terminating conditions for selected actions. Planning is predictive and functions in a progressive order. Execution of actions is allowed to begin when the first primitive behavior is reached. Any state has the ability to nullify successive steps in the plan, causing generation of a new plan to occur immediately (Chen et al., 2001 b). While this method of planning seems to conflict with the definition offered by Rich and Knight (1991), it is actually a small deviation that allows faster response times, more dynamic response to changing conditions, and the ability to take instruction from the user.

47

Missiaen proposes a planner using abduction (Missiaen, 1991). Like Chen et al., Missiaen's system is based on the Event Calculus, but Missiaen does not interweave perception, planning and action. His system uses a basic formalism in which the goal clause implies a set containing the assumptions that, along with the domain rules of the given problem, prove the goal clause. Localization of the problem domain is employed to decompose complex problems into critical sub-problems on which abductive planning is used. Missiaen goes on to describe an implementation of this system for planning the assembly of robots (Missiaen, 1991).

Abductive planning in decision making tasks has also been shown to be beneficial (Fox, 1999). Decision making in this context is defined as the process through which a character creates a plan of action to fulfill its needs. Fox proposes an algorithm that specifies a step-wise plan generator, evaluator, and assembler as a solution for a specific, but open-ended problem. To accomplish this, Fox employs a collection of information processing strategies that arrange potential plan steps into a taxonomical hierarchy, assess the plan steps using feature-based pattern matching and generate a statement of belief as to its relevance, and assemble the composite plan in a best-first manner, i.e., focus on a specific need and select the best plan-steps to fulfill that need. Abduction is employed during the classification of the potential plan-steps in order to find the best solution. In this case, the classification system generates a set of findings with respect to a group of hypotheses and abductive reasoning is used to derive an explanation of the findings that leads to the best composite (Fox, 1999).

Ajjanagadde (1993) proposes a system of 'generate and test' that is employed in many modern expert systems that incorporates background knowledge. Using the 'generate and test' method, backward chaining is used to generate a set of probable hypotheses, each of which are tested using the known facts to determine the most likely assumption. Ajjanagadde suggests that

48

this testing process should occur with some limited form of forward reasoning based on background knowledge. This distinction is important to processes such as planning (Ajjanagadde, 1993). Ajjanagadde's method is very similar to Klein's (1997) RPD model of human decision making.

Satoh and Yamamoto (2002) illustrate another instance in which agents (autonomous programs) are required to function with incomplete information. Given that agents must be able to communicate with other agents (Ferber, 1999) and that in many multi-agent systems communication failures caused by network latency or simply by the other agent's postponement of its decision, agents engaged in dialogue are often required to form beliefs in order to continue processing (Satoh and Yamamoto, 2002). In short, a belief is defined as an agent's default reasoning about some other agent or object (Satoh and Yamamoto, 2002). Satoh and Yamamoto (2002) propose a system that allows the agent to continue computation even when communication is not guaranteed. This system is based on a speculative belief computation using abduction (Satoh and Yamamoto, 2002).   Essentially, when a required response is not present, the system uses a default hypothesis as a response and continues computation. When the response is received, the speculative computation is verified and either validated or invalidated.

Satoh and Yamamoto's basic algorithm is as follows:

1. Each agent prepares a tentative answer for each possible question at the beginning of computation

2. When an agent asks a question of another, the agent uses its own tentative answer to continue computation as it waits for the response

3. When the response arrives, if the response validates the tentative answer, the agent continues processing. If, however, the response invalidates the tentative answer, the agent suspends computation, revises its belief, restarts processing based on the new belief, and updates its default answer to reflect the new answer.

In this instance, slow communications do not halt the system. While the information computed may not be accurate, as would be the case if a desired communiqué never appeared, but would have contained information contrary to the tentative answer, but this possibility always occurs when abduction is used. In a multi-agent system as Satoh and Yamamoto describe, the use of abduction implies that most computations at any given time may be labeled as tentative and subject to revision. From a philosophical point of view, this is acceptable as it closely models human communication (Satoh and Yamamoto, 2002).

Synthetic characters designed to display lexical and creative intentionality must be implemented with a computational cognitive model that supports automatacity, deliberate cognition, metacognition, domain-specific reasoning, decision making, planning and adaptive response, and learning and memory. While many excellent and well-formed theories have been presented, none support all the features mentioned above.

CHAPTER THREE: DESIGN AND METHODOLOGY

In the work, a model of personality that incorporated a trait-based taxonomy as expressed by the FFM and aspects of the psychological behaviorism theory as expressed by Staats was implemented. Additionally, a hierarchical model of mood and emotion with both dimensional and categorical features was implemented. Finally, a cognitive model that allows for automatic cognition with respect to natural language was implemented.

During the literature review presented in Chapter Two, it became clear to the author that in all three models required to create the work, two overwhelming similarities existed: (a) the psychological support for two distinct methods, a lexical method and a creative method, and (b) that the lexical method of each model was sufficient only for the implementation of lexical intentionality, but that the combination of lexical methods of one model with the lexical methods of the other two was sufficient for the implementation of creative intentionality as defined above. From the personality and emotive perspectives, Staats' psychological behaviorism relies on a lexicon of pattern-response pairs that are learned, and that can be combined or expanded to meet the requirements of new patterns. From the cognitive perspective, the concept of automatic responses to highly over-learned or practiced stimuli also relies on a lexicon of pattern-response pairs. Therefore, to implement lexical intentionality, any system that is able to utilize a pattern-response lexicon is sufficient. The implementation of creative intentionality was accomplished by adding a production system that allowed for the transformation of responses from the lexicon to meet the requirements of the current state. For example, a basic lexicon of pattern-response pairs was implemented to allow a synthetic character to respond appropriately to lexical atoms of human language. This system would satisfy the needs of lexical intentionality if it could

incorporate every ubiquitous word, phrase, and sentence in the target natural language and sensible constraints were placed on the input (i.e., not allowing the entire text of an encyclopedia to serve as a single input, etc.). Creative intentionality could be added to the above system by adding a set of production rules that govern composition and rare grammatical forms.

The synthetic personality implemented in the work operates in the domain of natural language. It implemented lexical communication and automatic cognition in the form of a Case-Based Reasoning (CBR) system. The CBR algorithm de-emphasizes the use of general rules, and instead stores problem cases and their solutions in long-term memory (Luger and Stubblefield, 1993). CBR is similar to analogical reasoning as discussed previously in that if the algorithm can't match the current problem to one its case library, it attempts to find a close match to the problem and transform that case's solution to fit.

To implement part of the cognitive model required for the work, a CBR system was implemented so that cases are defined by natural language strings of input and the responses take the form of natural language strings. This algorithm implemented lexical matching of natural language atoms and allows the synthetic character to display lexical intentionality and automatic cognition. It also fit with Staats' psychological behaviorism theory and filled the role of the linguistic-cognitive repertoires for the work.

To implement a basic personality, a model that combined both the tenets and structure of the FFM with a lexical and creative method for behavior selection was implemented. The synthetic personality was implemented by tracking five numeric values that correspond to an Extroversion scale, a Neurosis scale, an Agreeableness scale, an Openness scale, and a Conscientiousness scale. These scales ranged from 0.0 to 1.0, with $0.0 \leq x < 0.25$ denoting an abnormally low range, $0.26 \leq x \leq 0.9$ denoting a 'normal' range, $0.91 < x \leq 1.0$ denoting an

52

abnormally high range. Each of these scales impacted behavior selection, mood state traversal, and emotional state traversal as discussed below. In addition to these five values, each synthetic personality was defined with a disposition point that was skewed towards either a positive mood or a negative mood.

In the work, personalities with low values of the Neuroticism scale were expected to be emotionally fixed, while high values of the scale were expected to yield highly labile personalities. This was accomplished by a multiplicative increase to the signal strength of any input that is representational of the Neurosis factor. In other words, the signal strength of a given input was multiplied by the result of dividing the Neurosis factor by 0.5 (the neutral setting). In this manner, highly neurotic personality types more easily traversed to different emotional states, while personality types with a low Neurosis factor required higher signal strength inputs to successfully change state.

It was expected that a personality with a high Extroversion value appeared to have a more positive mood than a personality with a neutral setting. A personality with a low Extroversion value was expected to appear to suffer from a more negative mood. In the work, this was accomplished by changing the degree with which mood changes over time in response to the defined Extroversion value and by skewing the changes in either a positive or negative manner as dictated by the positive correlation between the Extroversion rating and mood.

A low Openness value in a personality was expected to be viewed as less emotionally intense. A high Openness value was expected to be seen as more emotionally intense than neutral or closed individuals. These criteria were met by implementing a multiplier applied to the signal strength when identifying the emotional intensity level.

Values within the high range of the Agreeableness scale were expected to cause personalities to have a more positive disposition and an initially higher mood value, while values within the low range of the Agreeableness scale began the simulation with a lower mood and a lower disposition setting.
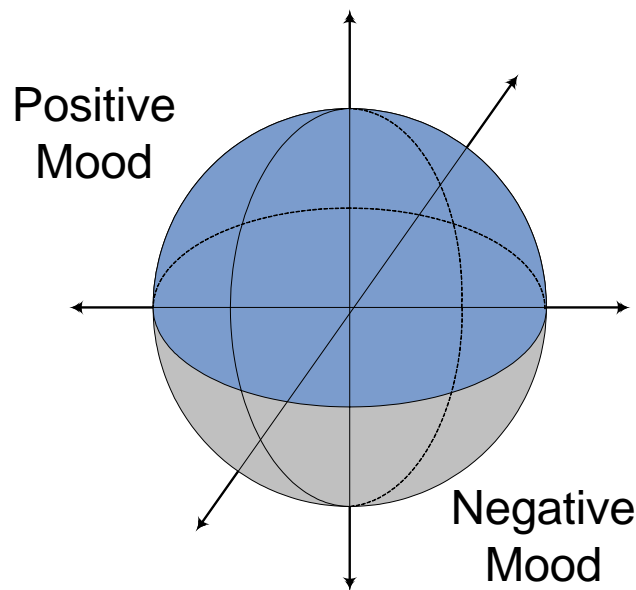
Figure 1 - Mood Model

The models of mood and emotion that were implemented can be explained by a beach ball metaphor with six slice-shaped sections and two hemispheres (see figures 1 and 2). Inputs to this model should be viewed as vectors in a three dimensional space (i.e., destination point and velocity), and thus emotional state transitions can be calculated by vector addition. These vectors are bounded by the skin of the sphere. The Openness factor mentioned above can be viewed as a multiplier that is applied to the velocity value of the vector. The top hemisphere will indicate positive mood, while the bottom hemisphere will indicate a negative mood. In this

metaphor, mood is calculated each time there is a change in emotional state by calculating the directional change to mood caused by the last emotional state.
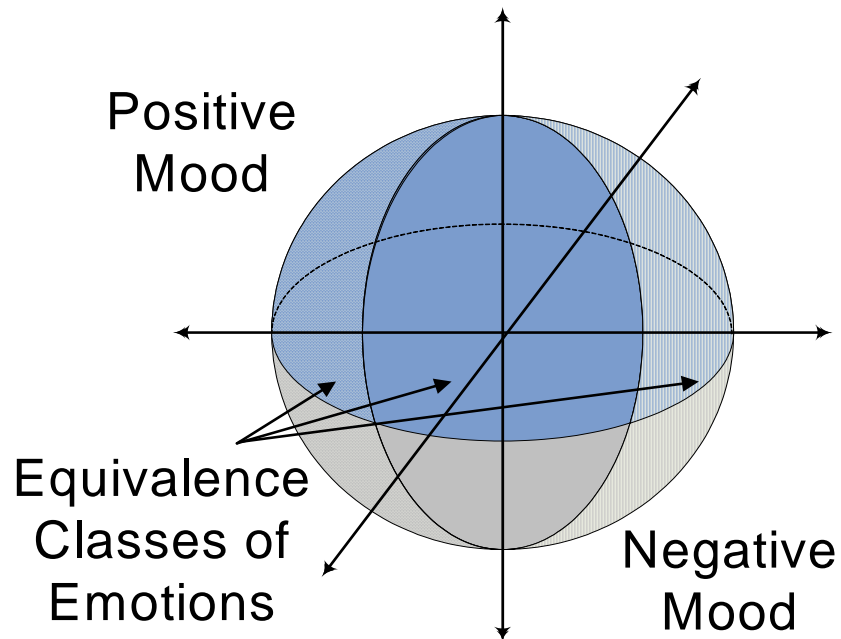
Figure 2 - Emotion Model

Each of the six slices represents a permutation on the set of human emotions, or an equivalence class of like emotions. In the work, the author labeled these equivalence classes: (a) joy, (b) anger, (c) sadness, (d) disgust, (e) surprise, and (f) fear, in accordance with Ekman's (1992) six basic emotions. The outer skin of the sphere represents the maximum intensity of emotion, and the origin of the sphere represents zero emotional intensity. An important feature of the representation is the varied width of the slice as one moves from the equator towards the poles of the sphere. This is important when one considers emotional lability – the farther from the origin of the sphere along the z-axis, the easier it becomes to change from one equivalence class to the next.

The Neurosis factor effectively dictates the point on the z-axis of the sphere (the height from the origin) at which the current emotional state calculations occur. In this manner, the width of the slice is narrowed, and the difference between zero intensity and maximum intensity is reduced as the point is moved away from the origin on the z-axis. As the calculation point approaches the skin of the sphere, the width of the slices approach zero, so that a highly labile character will be able to change equivalence classes with very little stimulus. As the point of calculation approaches the xy plane, the width of the slices increases to the maximum size, so that a highly stabile character will require significant stimulus to change its emotional state.

In addition to the CBR system that fulfilled the linguistic-cognitive repertoires of Staats' psychological behaviorism model, another CBR system was implemented that mapped the emotional-motivational repertoire of basic behaviors. In this system, each case represented stimulus/mental state-behavior pairs and a case-specific fitness function. Since the work was implemented in the domain of natural language, the stimulus portion was specified by a natural language string. The mental state portion of the case was specified in terms of the emotional state, and the intensity that successfully triggered the specified behavior. Behaviors ranged from transformations of natural language strings, creative expression in the form of natural language strings, mood state traversals, emotional state traversals, subtle changes to personality traits (in rare cases), or any combination of these processes.

<u>Hypotheses</u>

There are three methods of testing an implementation of a synthetic personality. The first is to verify that the software and the models are logically equivalent. That is to say, to test that

56

for all combinations of parameters, the software performs as predicted by the models. If this first test is successfully passed, anything that can be argued about the models is also true of the software and vice versa. The second is to verify the veracity of the models from a psychological standpoint. In other words, to ensure that the models behave as predicted by relevant psychological theory. Of course, the ultimate measure of the success of any synthetic personality is to test the believability of the responses generated by the system to human users. The last measure is outside the scope of this dissertation for two reasons: 1) in order for such a test to have validity, the synthetic personality would require significant content development (i.e. world knowledge, contextual knowledge), and 2) the systems presented in this dissertation are necessary, but not sufficient to create believable interactions with humans outside of iconic, innocuous exchanges. Specifically, the project would need increased NLU processing capability, the ability to intelligently modify phrases, sentences and paragraphs, increased cognitive abilities, and possibly perceptual and expressive systems capable of expression, gaze, and gesture.

The intent of this project was to create psychologically relevant implementations of synthetic personality, and synthetic personality dynamics, therefore the following hypotheses were considered:

Hypothesis One: The theoretical models of synthetic personality presented are logically equivalent to the implemented synthetic personality for all cases.

Hypothesis Two: The models of synthetic personality implemented in the planned work are true to the psychological theories on which they are based in all cases in which the effects of personality traits are predictable with certainty.
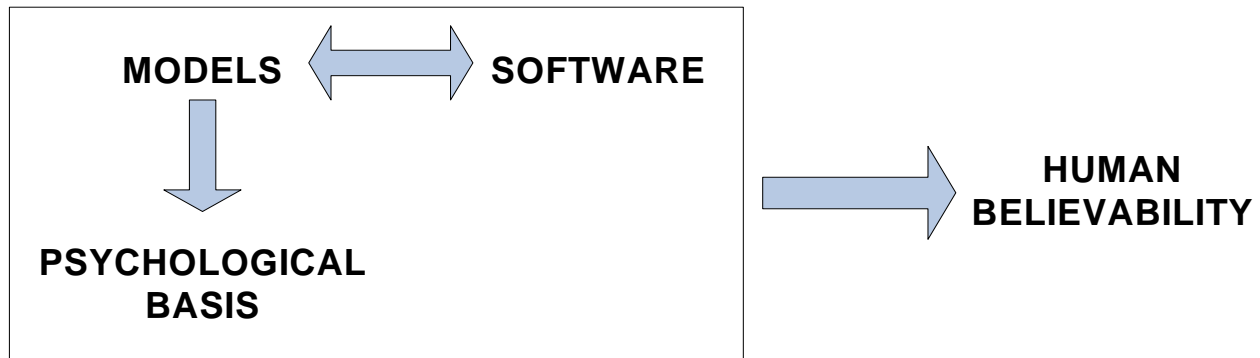
Figure 3 - Veracity of Synthetic Personalities

Evaluation Methods

Two methods of evaluation were used to verify the hypotheses. Both methods employed a synthetic personality and an independent machine observer that was be constructed and initialized with identical personality traits and emotion models. The observer processed an input through its version of the emotion model and personality traits, and recorded its estimates of the character's mental states. Simultaneously, the character being tested processed an identical input, logged its mental state, and responded via the user interface. Thus, at the end of any test, the logs contained both the actual mental state of the character based on the input, and the observer's estimation of what the mental state should be. In this manner, the logical equivalence between the models and the software was evaluated.

In order to accurately test the psychological veracity of the models, a systematic and complete approach was necessary. In some cases, specifically those cases in which a single trait

was high or low, and the other traits are neutral, an accurate prediction of behavior based on psychological theories and results was possible. However, a similar predictive evaluation of all possible combinations of traits and definitions was not possible, simply because complex interactions undoubtedly exist between multiple traits and not enough is known about these interactions in humans to predict the results in the models. Therefore, two distinct sets of test cases were called for and implemented. In each evaluation method, a specific set of test cases were run, as appropriate to the scope and intention of each method. In order to rule out chance, each test, in either method, was repeated twenty times. Since each test case was mutually exclusive from the other test cases, results from a given test case were analyzed and conclusions drawn against the particular parameters that define the test case and comparisons to baseline data were drawn.

A synthetic personality was defined with both permanently fixed and test case specific data. Fixed data included the definition of a weight table for the emotional state transitions, a weight table for the intensity state transitions, and a fixed set of CBR cases that defined a limited cognitive model for NLU. Test case specific data included definitions of individual traits, and individual mood and disposition settings.

The test cases described below each used the following weight tables:

Table 1 - Emotion Model Weight Table

State	Joy	Anger	Sadness	Disgust	Surprise	Fear
Joy	0.0	0.55	0.85	0.35	0.1	0.2
Anger	0.65	0.0	0.05	0.1	0.05	0.65
Sadness	0.85	0.15	0.0	0.05	0.25	0.35
Disgust	0.65	0.05	0.35	0.0	0.05	0.15
Surprise	0.05	0.05	0.05	0.05	0.0	0.05
Fear	0.45	0.15	0.05	0.05	0.05	0.0

Table 2 - Emotion State Intensity Weight Table

State - Intensity	1	2	3	4	5	6	7	8	9	10
Joy	0.1	0.1	0.1	0.2	0.2	0.2	0.3	0.3	0.3	0.3
Anger	0.1	0.1	0.1	0.1	0.2	0.3	0.4	0.5	0.5	0.6
Sadness	0.1	0.1	0.1	0.1	0.2	0.2	0.3	0.4	0.5	1.0
Disgust	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Surprise	0.1	0.1	0.1	0.1	0.2	0.3	0.3	0.4	0.5	0.6
Fear	0.5	0.5	0.6	0.6	0.7	0.7	0.8	0.8	0.9	1.0

For consistency between test cases, each test began with the emotion model artificially set to the joy state, and the mood and disposition equal to one another and set at negative, neutral, or positive (positively correlated with the Agreeableness trait). Personality trait definitions varied per test case, but within each case, were artificially fixed. Each test case was defined by the systematic evaluation of a set of inputs designed to test the transition between every state. Input evaluation included evaluation of the content and evaluation of the signal strength of the content, where the signal strength was required to exceed the weight associated with the input specified in the cognitive model for NLU.

To reduce the complexity of the evaluation methods, natural language inputs were simplified to single words, communicating both the nature of the input and the target emotion. For example, instead of adding a case to handle the phrase 'You look absolutely lovely!' a case was created to handle the emotional state the phrase was intended to be reduced to 'joy'. Therefore, the transitions used in testing were reduced to mapping a current emotional state, an input, and the signal strength of the input to an expected emotional state. CBR cases for each of the inputs were constructed as described below:

Table 3 - Inputs used for Evaluation

State	Input	Strength	Target	State	Input	Strength	Target
Joy	Joy	0.2	Joy	Disgust	Joy	0.7	Joy
Joy	Anger	0.6	Anger	Disgust	Anger	0.1	Anger
Joy	Sadness	0.9	Sadness	Disgust	Sadness	0.4	Sadness
Joy	Disgust	0.4	Disgust	Disgust	Disgust	0.1	Disgust
Joy	Surprise	0.1	Surprise	Disgust	Surprise	0.1	Surprise
Joy	Fear	0.2	Fear	Disgust	Fear	0.2	Fear
Joy	Joy	0.05	Joy	Disgust	Joy	0.6	Disgust
Joy	Anger	0.5	Joy	Disgust	Anger	0.025	Disgust
Joy	Sadness	0.8	Joy	Disgust	Sadness	0.3	Disgust
Joy	Disgust	0.3	Joy	Disgust	Disgust	0.05	Disgust
Joy	Surprise	0.05	Joy	Disgust	Surprise	0.025	Disgust
Joy	Fear	0.15	Joy	Disgust	Fear	0.1	Disgust
Anger	Joy	0.7	Joy	Surprise	Joy	0.1	Joy
Anger	Anger	0.5	Anger	Surprise	Anger	0.1	Anger
Anger	Sadness	0.1	Sadness	Surprise	Sadness	0.1	Sadness
Anger	Disgust	0.2	Disgust	Surprise	Disgust	0.1	Disgust
Anger	Surprise	0.1	Surprise	Surprise	Surprise	0.2	Surprise
Anger	Fear	0.7	Fear	Surprise	Fear	0.1	Fear
Anger	Joy	0.6	Anger	Surprise	Joy	0.025	Surprise
Anger	Anger	0.05	Anger	Surprise	Anger	0.025	Surprise
Anger	Sadness	0.025	Anger	Surprise	Sadness	0.025	Surprise
Anger	Disgust	0.5	Anger	Surprise	Disgust	0.025	Surprise
Anger	Surprise	0.025	Anger	Surprise	Surprise	0.05	Surprise
Anger	Fear	0.6	Anger	Surprise	Fear	0.025	Surprise
Sadness	Joy	0.9	Joy	Fear	Joy	0.5	Joy
Sadness	Anger	0.2	Anger	Fear	Anger	0.2	Anger
Sadness	Sadness	0.9	Sadness	Fear	Sadness	0.1	Sadness
Sadness	Disgust	0.1	Disgust	Fear	Disgust	0.1	Disgust
Sadness	Surprise	0.3	Surprise	Fear	Surprise	0.1	Surprise
Sadness	Fear	0.4	Fear	Fear	Fear	0.7	Fear
Sadness	Joy	0.8	Sadness	Fear	Joy	0.4	Fear
Sadness	Anger	0.1	Sadness	Fear	Anger	0.1	Fear
Sadness	Sadness	0.05	Sadness	Fear	Sadness	0.025	Fear
Sadness	Disgust	0.025	Sadness	Fear	Disgust	0.025	Fear
Sadness	Surprise	0.2	Sadness	Fear	Surprise	0.025	Fear
Sadness	Fear	0.3	Sadness	Fear	Fear	0.45	Fear

Test case specific trait definitions will be described with the definition of the test case below. In each test case, mood and disposition were initialized to the same value, 0.25, 0.5, or 0.75 for personalities with an Agreeableness factor of low, neutral, or high, respectively.

Evaluation Method One

In this method, baseline data and data about the isolated effects caused by manipulating single personality traits were acquired. Hypothesis One was evaluated by a comparison of all observer predictions of state and the actual states of the synthetic personality caused by evaluating the input. A match between the predicted state and the actual state for every input in each test case was identified as the metric for accepting Hypothesis One. Additionally, as isolated effects of high or low settings to a single personality trait on the emotion model could be predicted with certainty by the psychological theories upon which the models are based, Hypothesis Two was also tested by this method. That is to say, for each test case in method one, the mapping of input and current emotional state to the next emotional state, as specified above, was examined. It is acknowledged that the weights ascribed to the emotion model were somewhat arbitrary, but the same weight table was used in every test, and therefore the arbitrary definition of the weights did not impact the soundness of these measures. Additionally, for each test case, general trends of behavior were predicted based on relevant FFM research. Verification of the predicted trend by a manual inspection of the actual behavior produced by the system was performed. The razor for Hypothesis Two was identified as the matching of the actual behavioral trends of the synthetic personality and the trends predicted by psychological theory in every case.

In each of the test cases below, each of these inputs were artificially created and passed into the character. The resulting emotion was verified against these predicted values and discrepancies were noted. Additionally, the inputs were passed into the observer, and the internal validity of the computational models was verified as described above.

Test Case One: As a baseline, the system was tested against a personality defined by "neutral" settings for all of the five personality traits. This test case also provided "neutral" emotion results for comparison in future evaluations.

Table 4 - Trait Definitions for Test Case One

Trait	TC1
O	N
C	N
E	N
A	N
N	N

In this test case, psychological evidence predicted that the synthetic personality would react exactly as specified according to the emotion weight table, the intensity weight table, and mood would be tightly coupled with disposition and emotional state. In other words, the emotion model and the intensity model would be followed precisely, and mood would follow the general trend of emotion (positively correlated), returning to the disposition in the absence of changes to the emotion system over time.

For example, if the synthetic personality was in the Joy state, and receives an input of Anger with a signal strength of 0.6, it was expected that the synthetic personality would traverse to the Anger state, with low intensity (0), and its mood would begin to decline from the present

level towards the negative axis. Similarly, if the next input was Anger with a signal strength of 0.1, it was predicted that the synthetic personality would remain in the Anger state with a moderate intensity (4), and its mood would continue to decline towards the negative axis.

Test Cases Two through Six: In these test cases, each of the five personality traits were elevated to the high level in turn, while the other personality traits remained at the neutral setting (e.g., in test case four, Extroversion were set at the high level, while Neurosis, Agreeableness, Conscientiousness, and Openness were set to the neutral level). These cases tested changes in the emotion and mood models caused by elevation of one personality trait.

Table 5 - Trait Definitions for Test Cases Two through Six

Trait	TC2	TC3	TC4	TC5	TC6
O	H	N	N	N	N
C	N	H	N	N	N
E	N	N	H	N	N
A	N	N	N	H	N
N	N	N	N	N	H

In Test Case Two, since the Openness factor is set to high, it was predicted by psychological theory that the synthetic personality would react exactly as specified by the emotion weight table, and that mood would be positively related to disposition and emotional state. The intensity of the emotion would be elevated with respect to the defined intensity table. That is to say, for the inputs that signaled a change in intensity in the baseline data, a bigger change in intensity would be present.

In other words, the synthetic personality defined by Test Case Two would be seen as more intensely emotional. By contrast, it would not move between emotional states more easily

as would be the case if Neurosis was elevated, but would experience more intensity of emotion in any given state than the baseline data (or any other personality defined with a neutral Openness factor). To be clear, signal strength used for calculating the intensity level of emotion for personalities defined in this test case was multiplied by a value equal to the Openness factor divided by the neutral definition of the Openness factor (in this case, 0.95 / 0.5 = 1.9).

In Test Case Three, since the Conscientiousness factor was set to high, it was predicted that the synthetic personality would react exactly as specified according to the emotion weight table, the intensity weight table, and mood would be tightly coupled with disposition and emotional state. In other words, this test case would exactly mirror the baseline data since FFM theory shows no correlation between Conscientiousness and emotion, intensity, or mood.

In Test Case Four, since relevant psychological research indicated that the Extroversion rating is positively correlated to mood, and since Extroversion was set to high in this case, the synthetic personality was expected to react exactly as specified according to the emotion weight table, and the intensity weight table, but mood would not be tightly coupled with disposition and emotional state. Instead, mood would be more positive with respect to the baseline data.

In other words, while emotional state traversal and intensity state traversal would not be different from the baseline data, mood would tend to be more positive in scale, and positive more of the time than in the baseline data. Additionally, negative moves in mood would be smaller than in the baseline data. For example, if the baseline data showed that a change from the Joy state to the Anger state should lead to a decrease in the mood value of 0.05, the synthetic personality might have shown the same traversal in emotion but a decrease in mood of only 0.02 or a similar value. Likewise, while baseline data might have shown that a traversal from the Anger state to the Joy state resulted in an increase in mood of 0.05, the synthetic personality

might have shown the same traversal in emotion but an increase in mood of 0.10 or a similar value. In this manner, since the negative changes in mood were smaller, the synthetic personality remained in a positive mood longer than a personality with a neutral Extroversion rating, and when the mood dipped into the negative range, it would not dip as far and would regain positivity more easily.

In Test Case Five, since Agreeableness was elevated to the high level, psychological research indicated that the synthetic personality would react exactly as specified by the intensity weight table, and that mood would be tightly coupled with disposition and emotional state. The emotional state, however, was expected to show a higher incidence of positive axis emotion. That is to say, for the inputs that signaled a transition to a negative axis emotion may have resulted in no transition from the joy state. This was accomplished by increasing the signal strength of positive axis inputs, and decreasing the signal strength of negative axis inputs.

For example, if an input was associated with a traversal to the Joy state and was given a signal strength of 0.04, and the baseline personality was in the Surprise state, no traversal would occur, given the weight tables described above. If the same input was provided to the synthetic personality in this test case, the signal strength would be multiplied by 125% and a transition to the Joy state would occur. Similarly if an input was associated with a traversal to the Anger state, and was given a signal strength of 0.044, and the baseline personality is in the Joy state, no traversal to the Anger state would occur. If the same input was provided to the personality defined in this case, however, the signal strength would be multiplied by 125% and a traversal would occur.

In Test Case Six, a highly Neurotic personality was created and in accordance with FFM theory, it was predicted that the synthetic personality's mood would be tightly coupled with

disposition and emotional state. The emotional state or the intensity state, however, would transition to other emotion or intensity states with signal strengths that were lower than required in the baseline data. That is to say, for the inputs that failed to signal a transition to an alternate emotion or intensity in the baseline data, transitions may have occurred.

For example, if an input was created that signals a traversal to the Joy state and was given a signal strength of 0.4, and the baseline personality was in the Disgust state, no traversal would occur. If, however, this input was passed into the personality defined in this case, the signal strength was multiplied by a value equal to the Neurosis factor divided by the neutral definition of the Neurosis factor (in this case, 0.95 / 0.5 = 1.9), which would result in a sufficient signal strength to trigger a traversal to the Joy state. Likewise, if the input signaled a traversal to the Joy state and was given a signal strength of 0.15, and the baseline personality was in the Joy state, the resulting intensity level would be 3. If this input was passed into the personality defined by this case, the signal strength is again modified by the method described above, which would result in an intensity value of 7.

Test Cases Seven through Eleven: In these test cases, each of the five personality traits was lowered to the low level in turn, while the other personality traits remained at the neutral setting (e.g., in test case nine, Extroversion was set at the low level, while Neurosis, Agreeableness, Conscientiousness, and Openness were all set to the neutral level). These cases tested changes in the emotion model caused by a low setting of one personality trait.

Table 6 - Trait Definitions for Test Cases Seven through Eleven

Trait	TC7	TC8	TC9	TC10	TC11
O	L	N	N	N	N
C	N	L	N	N	N
E	N	N	L	N	N
A	N	N	N	L	N
N	N	N	N	N	L

In Test Case Seven, since the Openness factor was set to low, it was predicted by psychological theory that the synthetic personality would react exactly as specified by the emotion weight table, and that mood would be positively related to disposition and emotional state. The intensity of the emotion, however, would be lower with respect to the defined intensity table. That is to say, for the inputs that signaled a change in intensity in the baseline data, a smaller change, or no change at all in intensity would be present.

In other words, the synthetic personality defined by Test Case Seven would be seen as less intensely emotional. By contrast, it would appear to be flatter in affect as would be the case if Neurosis was lowered, but would experience less intensity of emotion in any given state than the baseline data (or any other personality defined with a neutral Openness factor). To be clear, signal strength used for calculating the intensity level of emotion for personalities defined in this test case was multiplied by a value equal to the Openness factor divided by the neutral definition of the Openness factor (in this case, $0.25 / 0.5 = 0.5$).

In Test Case Eight, since the Conscientiousness factor was set to low, it was predicted that the synthetic personality would react exactly as specified according to the emotion weight table, the intensity weight table, and mood would be tightly coupled with disposition and

emotional state. In other words, this test case would exactly mirror the baseline data since FFM theory shows no correlation between Conscientiousness and emotion, intensity, or mood.

In Test Case Nine, since relevant psychological research indicated that the Extroversion rating is positively correlated to mood, and since Extroversion was set to low in this case, the synthetic personality was expected to react exactly as specified according to the emotion weight table, and the intensity weight table, but mood was not expected to be tightly coupled with disposition and emotional state. Instead, mood was expected to be more negative with respect to the baseline data.

In other words, while emotional state traversal and intensity state traversal would not be different from the baseline data, mood would tend to be more negative in scale, and negative more often than in the baseline data. Additionally, positive moves in mood would be smaller than in the baseline data. For example, if the baseline data showed that a change from the Joy state to the Anger state should lead to a decrease in the mood value of 0.05, the synthetic personality might have shown the same traversal in emotion but a decrease in mood of 0.10 or a similar value. Likewise, while baseline data might have shown that a traversal from the Anger state to the Joy state resulted in an increase in mood of 0.05, the synthetic personality might have shown the same traversal in emotion but an increase in mood of 0.02 or similar value. In this manner, since the positive changes in mood were smaller, the synthetic personality would remain in a negative mood longer than a personality with a neutral Extroversion rating, and when the mood traversed into the positive range, would not go as far and would regain negativity more easily.

In Test Case Ten, since Agreeableness was low, psychological research indicated that the synthetic personality would react exactly as specified by the intensity weight table, and that

mood would be tightly coupled with disposition and emotional state. The emotional state, however, was expected to show a higher incidence of negative axis emotion. That is to say, for the inputs that signaled a transition to a positive axis emotion no transition from a negative axis state would result. This was accomplished by decreasing the signal strength of positive axis inputs, and increasing the signal strength of negative axis inputs.

For example, if an input was associated with a traversal to the Joy state and was given a signal strength of 0.05, and the baseline personality was in the Surprise state, a traversal would occur, given the weight tables described above. If the same input was provided to the synthetic personality in this test case, the signal strength would be multiplied by 65% and no transition to the Joy state would occur. Similarly if an input was associated with a traversal to the Anger state, and was given a signal strength of 0.085, and the baseline personality was in the Joy state, a traversal to the Anger state would occur. If the same input was provided to the personality defined in this case, however, the signal strength would be multiplied by 65% and no traversal would occur.

In Test Case Eleven, personality with a low Neurosis factor was created and in accordance with FFM theory, it was predicted that the synthetic personality's mood would be tightly coupled with disposition and emotional state. The emotional state or the intensity state, however, was expected to transition to other emotion or intensity states only at signal strengths that were much higher than required in the baseline data. That is to say, for the inputs that signaled a transition to an alternate emotion or intensity in the baseline data, transitions may not have occurred.

For example, if an input was created that signals a traversal to the Joy state and was given a signal strength of 0.75, and the baseline personality was in the Disgust state, a traversal to the

Joy state would occur. If, however, this input was passed into the personality defined in this case, the signal strength was multiplied by a value equal to the Neurosis factor divided by the neutral definition of the Neurosis factor (in this case, 0.25 / 0.5 = 0.5), which would result in insufficient signal strength to trigger a traversal to the Joy state. Likewise, if the input signaled a traversal to the Joy state and was given a signal strength of 0.15, and the baseline personality was in the Joy state, the resulting intensity level would be 3. If this input was passed into the personality defined by this case, the signal strength was again modified by the method described above, which would result in an intensity value of 0.

Evaluation Method Two

In this method, the interactions between personality traits were too complex or too dependent to be adequately predicted. As such, test cases in this method could not be used to satisfy Hypothesis Two and should be viewed as exploratory tests of the complex interactions between personality traits. Since these test cases attempted to verify that the complex interactions of the personality modeling do not adversely effect the proper functioning of the emotion and mood modeling, the cases specified below were used to satisfy Hypothesis One. To that end, the same set of inputs was run as in method one, but verification that the emotion model actually transitions to the target state was not performed. Additionally, due to the complexity and unknown nature of the interactions between multiple personality traits, no predictions of general trends in behavior could be made. Instead, the data generated in this method was used to verify that the personality model did not impact the internal validity of the computational models as indicated by the ability of the independent observer to accurately estimate the mental state of

71

the character. Current state, input, signal strength and resulting state data were recorded, however, for use in future work.

Test Cases Twelve through Twenty-One: In these test cases, the interaction between two high values, and three neutral values of personality traits was tested (e.g. Extroversion and Neurosis were set to high, while Agreeableness, Conscientiousness, Openness were set to neutral).

Table 7 - Trait Definitions for Test Cases Twelve through Twenty-One

Trait	TC12	TC13	TC14	TC15	TC16	TC17	TC18	TC19	TC20	TC21
O	H	H	H	H	N	N	N	N	N	N
C	H	N	N	N	H	H	H	N	N	N
E	N	H	N	N	H	N	N	H	H	N
A	N	N	H	N	N	H	N	H	N	H
N	N	N	N	H	N	N	H	N	H	H

Test Cases Twenty-Two through Thirty-One: In these test cases, the interaction between two low values, and three neutral values of personality traits was tested (e.g. Extroversion and Neurosis were set to low, while Agreeableness, Conscientiousness, Openness were set to neutral).

Table 8 - Trait Definitions for Test Cases Twenty-two through Thirty-One

Trait	TC22	TC23	TC24	TC25	TC26	TC27	TC28	TC29	TC30	TC31
O	L	L	L	L	N	N	N	N	N	N
C	L	N	N	N	L	L	L	N	N	N
E	N	L	N	N	L	N	N	L	L	N
A	N	N	L	N	N	L	N	L	N	L
N	N	N	N	L	N	N	L	N	L	L

Test Cases Thirty-Two through Forty-One: These test cases tested the effect of three high traits and two neutral traits (e.g. Openness, Agreeableness, and Conscientiousness were set to high while Extroversion and Neurosis were set to neutral).

Table 9 - Trait Definitions for Test Cases Thirty-Two through Forty-One

Trait	TC32	TC33	TC34	TC35	TC36	TC37	TC38	TC39	TC40	TC41
O	H	H	H	H	H	H	N	N	N	N
C	H	H	H	N	N	N	H	H	H	N
E	H	N	N	H	H	N	H	H	N	H
A	N	H	N	H	N	H	H	N	H	H
N	N	N	H	N	H	H	N	H	H	H

Test Cases Forty-Two through Fifty-One: These test cases tested the effect of three low traits and two neutral traits (e.g. Openness, Agreeableness, and Conscientiousness were set to high while Extroversion and Neurosis were set to neutral).

Table 10 - Trait Definitions for Test Cases Forty-Two through Fifty-One

Trait	TC42	TC43	TC44	TC45	TC46	TC47	TC48	TC49	TC50	TC51
O	L	L	L	L	L	L	N	N	N	N
C	L	L	L	N	N	N	L	L	L	N
E	L	N	N	L	L	N	L	L	N	L
A	N	L	N	L	N	L	L	N	L	L
N	N	N	L	N	L	L	N	L	L	L

Test Cases Fifty-Two through Fifty-Six: These test cases tested the effect of four high traits

and two neutral traits (e.g. Openness, Agreeableness, Conscientiousness, and Extroversion were

set to high while Neurosis was set to neutral).

Table 11 - Trait Definitions for Test Cases Fifty-Two through Fifty-Six

Trait	TC52	TC53	TC54	TC55	TC56
O	H	H	H	H	N
C	H	H	H	N	H
E	H	H	N	H	H
A	H	N	H	H	H
N	N	H	H	H	H

Test Cases Fifty-Seven through Sixty-One: These test cases tested the effect of four low traits

and two neutral traits (e.g. Openness, Agreeableness, Conscientiousness, and Extroversion were

set to low while Neurosis was set to neutral).

Table 12 - Trait Definitions for Test Cases Fifty-Seven through Sixty-One

Trait	TC57	TC58	TC59	TC60	TC61
O	L	L	L	L	N
C	L	L	L	N	L
E	L	L	N	L	L
A	L	N	L	L	L
N	N	L	L	L	L

Test Case Sixty-Two: This test case simulated the effect of five high personality traits. This

test case also provided "high" emotion results for comparison in future evaluations.

Table 13 - Trait Definitions for Test Case Sixty-Two

Trait	TC62
O	H
C	H
E	H
A	H
N	H

Test Cases Sixty-Three through Sixty-Seven: In these test cases, each of the five personality

traits was set at a low level in turn, while the other personality traits were set to high (e.g., in test

case sixty-five, Extroversion was set at the low level, while Neurosis, Agreeableness,

Conscientiousness, and Openness were set to the high level).

Table 14 - Trait Definitions for Test Cases Sixty-Three through Sixty-Seven

Trait	TC63	TC64	TC65	TC66	TC67
O	L	H	H	H	H
C	H	L	H	H	H
E	H	H	L	H	H
A	H	H	H	L	H
N	H	H	H	H	L

Test Cases Sixty-Eight through Seventy-Seven: In these test cases, the interaction between two low values, and three high values of personality traits was tested (e.g. Extroversion and Neurosis were set to low, while Agreeableness, Conscientiousness, Openness were set to high).

Table 15 - Trait Definitions for Test Cases Sixty-Eight through Seventy-Seven

Trait	TC68	TC69	TC70	TC71	TC72	TC73	TC74	TC75	TC76	TC77
O	H	H	H	H	H	H	L	L	L	L
C	H	H	H	L	L	L	H	H	H	L
E	H	L	L	H	H	L	H	H	L	H
A	L	H	L	H	L	H	H	L	H	H
N	L	L	H	L	H	H	L	H	H	H

Test Cases Seventy-Eight through Eighty-Seven: These test cases tested the effect of three low traits and two neutral traits (e.g. Openness, Agreeableness, and Conscientiousness were set to high while Extroversion and Neurosis were set to low).

Table 16 - Trait Definitions for Test Cases Seventy-Eight through Eighty-Seven

Trait	TC78	TC79	TC80	TC81	TC82	TC83	TC84	TC85	TC86	TC87
O	L	L	L	L	L	L	H	H	H	H
C	L	L	L	H	H	H	L	L	L	H
E	L	H	H	L	L	H	L	L	H	L
A	H	L	H	L	H	L	L	H	L	L
N	H	H	L	H	L	L	H	L	L	L

Test Cases Eighty-Eight through Ninety-two: These test cases tested the effect of four low

traits and one high traits (e.g. Openness, Agreeableness, Conscientiousness, and Extroversion

were all set to low while Neurosis was set to high).

Table 17 - Trait Definitions for Test Cases Eighty-Eight through Ninety-Two

Trait	TC88	TC89	TC90	TC91	TC92
O	L	L	L	L	H
C	L	L	L	H	L
E	L	L	H	L	L
A	L	H	L	L	L
N	H	L	L	L	L

Test Case Ninety-Three: This test case simulated the effect of five low personality traits. This

test case also provided "low" emotion results for comparison in future evaluations.

Table 18 - Trait Definitions for Test Case Ninety-Three

Trait	TC93
O	L
C	L
E	L
A	L
N	L

Test Cases Ninety-Four through One Hundred Thirteen: In these test cases, the interaction between a single high value, a single low value, and three neutral values of personality traits was tested (e.g. Extroversion was be set to high, Neurosis was set to low, while Agreeableness, Conscientiousness, Openness were set to neutral).

Table 19 - Trait Definitions for Test Cases Ninety-Four through One Hundred Thirteen

Trait	TC94	TC95	TC96	TC97	TC98	TC99	TC100	TC101	TC102	TC103
O	H	H	H	H	L	L	L	L	N	N
C	L	N	N	N	H	N	N	N	H	H
E	N	L	N	N	N	H	N	N	L	N
A	N	N	L	N	N	N	H	N	N	L
N	N	N	N	L	N	N	N	H	N	N

Trait	TC104	TC105	TC106	TC107	TC108	TC109	TC110	TC111	TC112	TC113
O	N	N	N	N	N	N	N	N	N	N
C	H	L	L	L	N	N	N	N	N	N
E	N	H	N	N	H	H	L	L	N	N
A	N	N	H	N	L	N	H	N	H	L
N	L	N	N	H	N	L	N	H	L	H

Test Cases One Hundred Fourteen through One Hundred Forty-Three: In these test cases, the interaction between a two low values, a single high value, and two neutral values of personality traits was tested (e.g. Extroversion and Openness were set to low, Agreeableness was set to high, while Neurosis, Conscientiousness were set to neutral).

Table 20 - Trait Definitions for Test Cases One Hundred Fourteen through One Hundred Forty-Three

Trait	TC114	TC115	TC116	TC117	TC118	TC119	TC120	TC121	TC122	TC123	TC124
O	H	H	H	H	H	H	N	N	N	L	L
C	L	L	L	N	N	N	H	H	H	H	H
E	L	N	N	L	L	N	L	L	N	L	N
A	N	L	N	L	N	L	L	N	L	N	L
N	N	N	L	N	L	L	N	L	L	N	N
Trait	TC125	TC126	TC127	TC128	TC129	TC130	TC131	TC132	TC133	TC134	TC135
O	L	N	L	N	L	N	L	N	N	L	L
C	H	N	N	L	N	L	L	L	L	N	L
E	N	H	H	H	H	H	H	L	N	N	N
A	N	L	L	L	N	N	N	H	H	H	H
N	L	L	N	N	L	L	N	N	L	L	N

Trait	TC136	TC137	TC138	TC139	TC140	TC141	TC142	TC143
O	N	L	N	L	L	N	L	N
C	N	N	L	L	N	L	N	N
E	L	L	L	N	N	N	L	L
A	H	H	N	N	L	L	N	L
N	L	N	H	H	H	H	H	H

Test Cases One Hundred Forty-Four through One Hundred Seventy-Three: In these test cases, the interaction between a two high values, a single low value, and two neutral values of personality traits was tested (e.g. Extroversion and Openness were set to high, Agreeableness was set to low, while Neurosis, Conscientiousness was set to neutral).

Table 21 - Trait Definitions for Test Cases One Hundred Fourty-Four through One Hundred Seventy-Three

Trait	TC144	TC145	TC146	TC147	TC148	TC149	TC150	TC151	TC152	TC153	TC154
O	H	H	H	H	H	H	N	N	N	L	L
C	L	L	L	N	N	N	H	H	H	H	H
E	L	N	N	L	L	N	L	L	N	L	N
A	N	L	N	L	N	L	L	N	L	N	L
N	N	N	L	N	L	L	N	L	L	N	N
Trait	TC155	TC156	TC157	TC158	TC159	TC160	TC161	TC162	TC163	TC164	TC165
O	L	N	L	N	L	N	L	N	N	L	L
C	H	N	N	L	N	L	L	L	L	N	L
E	N	H	H	H	H	H	H	L	N	N	N
A	N	L	L	L	N	N	N	H	H	H	H
N	L	L	N	N	L	L	N	N	L	L	L

Trait	TC166	TC167	TC168	TC169	TC170	TC171	TC172	TC173
O	N	L	N	L	L	N	L	N
C	N	N	L	L	N	L	N	N
E	L	L	L	N	N	N	L	L
A	H	H	N	N	L	L	N	L
N	L	N	H	H	H	H	H	H

Test Cases One Hundred Seventy-Four through One Hundred Ninety-Three: In these test cases, the interaction between a three low values, a single high value, and one neutral values of personality traits was tested (e.g. Extroversion, Agreeableness, and Openness were set to low, Conscientiousness was set to high, while Neurosis was set to neutral).

Table 22 - Trait Definitions for Test Cases One Hundred Seventy-Four through One Hundred Ninety-Three

Trait	TC174	TC175	TC176	TC177	TC178	TC179	TC180	TC181	TC182	TC183
O	H	H	H	H	N	N	N	N	L	L
C	L	L	L	N	L	L	L	H	L	L
E	L	L	N	L	L	L	H	L	L	H
A	L	N	L	L	L	H	L	L	H	L
N	N	L	L	L	H	L	L	L	N	N
Trait	TC184	TC185	TC186	TC187	TC188	TC189	TC190	TC191	TC192	TC193
O	L	L	L	L	L	L	L	L	L	L
C	H	L	L	N	H	H	N	N	L	L
E	L	L	N	L	L	N	L	H	H	N
A	L	N	L	L	N	L	H	L	N	H
N	N	H	H	H	L	L	L	L	L	L

Test Cases One Hundred Ninety-Four through Two Hundred Eleven: In these test cases,

the interaction between a three high values, a single low value, and one neutral values of

personality traits was tested (e.g. Extroversion, Agreeableness, and Openness were set to high,

Conscientiousness was set to low, while Neurosis was set to neutral).

Table 23 - Trait Definitions for Test Cases One Hundred Ninety-Four through Two Hundred

Eleven

Trait	TC194	TC195	TC196	TC197	TC198	TC199	TC198	TC199	TC200	TC201
O	L	L	L	L	N	N	N	N	H	H
C	H	H	H	N	H	H	H	L	H	H
E	H	H	N	H	H	H	L	H	H	L
A	H	N	H	H	H	L	H	H	L	H
N	N	H	H	H	L	H	H	H	N	N
Trait	TC202	TC203	TC204	TC205	TC206	TC207	TC208	TC209	TC210	TC211
O	H	H	H	H	H	H	H	H	H	H
C	L	H	H	N	L	L	N	N	H	H
E	H	H	N	H	H	N	H	L	L	N
A	H	N	H	H	N	H	L	H	N	L
N	N	L	L	L	H	H	H	H	H	H

By exhaustively testing every case in the problem domain, and by comparing the results

of these cases against one another, the author expected to be able to identify common trends and

errors in the computational models. Additionally, the results from this exhaustive testing could

be used in future works as baseline data or as comparison points for data generated by human

testing.

CHAPTER FOUR: RESULTS AND CONCLUSIONS

The synthetic personality models proposed above were implemented and tested as specified in Chapter Three. Data from the entire problem domain was collated and examined, and when viewed from a trait definition perspective (e.g. considering all cases with a high Extroversion and low Neurosis, etc.), the following general trends emerged:

1. High Agreeableness has more impact on emotional response that does any Neurosis value.

2. Emotional response mirrors the baseline data if and only if: either both Agreeableness and Neurosis are neutral or both Openness and Neurosis are neutral while Extroversion is not high.

3. In terms of intensity response, Openness has more impact than Neurosis.

4. In order for intensity response to mirror the baseline data, both Openness and Neurosis must be neutral.

5. Agreeableness can override the impact of high Extroversion in terms of mood response if and only if Neurosis is neutral.

6. Extremely flat emotional response (large decrease from baseline data) has more impact on mood than does low or neutral Extroversion.

7. Mood response and average change in mood show a mild decrease if and only if Extroversion is low and Openness, Agreeableness, and Neurosis are neutral.

8. If Extroversion is neutral, Agreeableness is low and emotions are not flat, mood response shows a small decrease, while the average change in mood shows a moderate decrease.

9. Mood response and average change in mood mirror the baseline data if and only if Extroversion, Agreeableness, and Neurosis are neutral.

A total of 4220 test runs were made (twenty runs for each test case), and of those 4220 test runs, the independent observer was able to predict the exact mental state of the synthetic personality for every input in every run. Given the 100% rate of success, it is concluded that the theoretical models of synthetic personality presented are logically equivalent to the implemented synthetic personality for all cases. Therefore, Hypothesis One is accepted.

The data gathered as a result of evaluation method one is summarized in the following table. The column labeled TC reflects the test case number for the row. The columns labeled O, C, E, A, N reflect the Openness, Conscientiousness, Extroversion, Agreeableness, and Neurosis trait values used in the test case, with L, N, H indicating a low, neutral or high setting. The columns labeled ER, IR, MR indicate the emotional result, intensity result, and mood result. These columns are filled with -, +, BL, or a * to indicate that the result is below baseline, above baseline, the same as baseline, or different from the baseline data, but as expected, respectively. Note that in some cases, the model acts as expected but produces different values than the baseline data. For example, a high agreeableness setting will result in more positive axis emotions, and so it is expected that mood would show a higher incidence of positivity and that more intensity will be seen (as inputs won't signal a change to an alternate state but might cause intensity state transitions).

Table 24 - Emotional, Intensity, and Mood Response Results for Test Cases One through Eleven

TC	O	C	E	A	N	ER	IR	MR
1	N	N	N	N	N	BL	BL	BL
2	H	N	N	N	N	BL	+	BL
3	N	H	N	N	N	BL	BL	BL
4	N	N	H	N	N	BL	BL	+
5	N	N	N	H	N	+	*	*
6	N	N	N	N	H	+	+	*
7	L	N	N	N	N	BL	-	BL
8	N	L	N	N	N	BL	BL	BL
9	N	N	L	N	N	BL	BL	-
10	N	N	N	L	N	-	*	*
11	N	N	N	N	L	-	-	*

In Test Case Two, the emotional response and mood response exactly mirrored the baseline data. The intensity response, however, was elevated. For example, where an input signaled a transition to an intensity level of 5 in the baseline data, an intensity level of 9 was observed in Test Case Two. While this large change in intensity was not uniform due to different intensity weights for each emotion state, the overall elevation was uniform (i.e. every transition that signals a change in intensity in the baseline data signals a change to a higher intensity state in this test case).

Psychological theory dictates that individuals with a high Openness factor should mirror the baseline data in terms of emotion and mood, but should be more intensely emotional. Since the data indicates this is true for Test Case Two, this test supports Hypothesis Two.

In Test Case Three, no difference in emotional, intensity, or mood response as compared to the baseline data was observed. Since psychological evidence shows no correlation between Conscientiousness and emotion, intensity, or mood, Test Case Three supports Hypothesis Two.

In Test Case Four, emotion and intensity was indistinguishable from the baseline data, while mood was shown to be markedly more positive. In fact, the mood rating never dipped below the median point in the mood scale, and after the first two inputs in the sequence, showed an average value of 0.78 on a 0.0 to 1.0 scale. By way of comparison, the baseline data average for the same range was 0.09. In addition, when tested on four sequential inputs that evaluate to the joy state, the baseline case showed an increase of 0.2 in mood, while Test Case Four shows an increase of 0.3. When faced with five sequential inputs that evaluate to anger, the baseline case showed a decrease in mood of 0.35, whereas this test case showed a decrease of only 0.04.

Since psychological evidence indicates that individuals with a high Extroversion factor are expected to have a more positive mood, this test case supports Hypothesis Two.

In Test Case Five, both intensity and mood response was as expected, although substantially different from the baseline data. This difference is due to the expected changes within the model caused by the change from normal emotion state transitions and the transitions induced by a high Agreeableness factor. Specifically, the high Agreeableness factor caused the emotion response for this test case to become 'trapped' in the joy state. Since every input resulted in a self-loop transition on the joy state, an intensity transition was triggered for every input. Similarly, since the mood model is impacted by the emotion state, and since the emotion state for this test case was always a positive one, mood was positive for every input as well.

FFM theory predicts that individuals with a high Agreeableness factor show a higher incidence of positive emotion. The data from Test Case Five clearly supports this prediction and therefore, this test case supports Hypothesis Two.

In Test Case Six, both emotion response and intensity response was elevated with respect to the baseline data. Specifically, in response to inputs that did not trigger a transition to another

86

emotion state in the baseline case, this test case changed state every time. Also, as in Test Case Two, the overall elevation of intensity was uniform. Mood response was as expected but different from that of the baseline run. As in the previous case, this is a result of the change in emotion state.

Evidence from clinical trial shows that highly Neurotic individuals are more emotionally labile, both in terms of escalation of intensity and in terms of changes in emotion state. Test Case Six supports Hypothesis Two.

In Test Case Seven, the intensity response was decreased, while emotion and mood response was identical to the baseline case. For example, an input that caused an intensity change to 8 in the baseline data only signaled a change to an intensity level of 3 in this test case. As in Test Case Two, the amount of change was dependent on the emotion state, but the uniform decrease in response was seen.

Since Openness is correlated with emotional intensity in psychology research, Test Case Seven supports Hypothesis Two.

In Test Case Eight, no difference between the baseline data and data derived from this test case was observed. Since there is no correlation between Conscientiousness and emotion, intensity or mood in the FFM theory, this test case supports Hypothesis Two.

In Test Case Nine, both the emotion response and the intensity response mirror that found in the baseline case. Mood response was found to be lower than that of the baseline data. Specifically, the average mood of the character after it moved away from its disposition setting was 0.04 in this test case, as opposed to 0.09 in the baseline data. Also, in the four inputs designed to maintain the joy state, the baseline case changed 0.2 mood points. In this case, however, the same inputs resulted in a change of 0.035. Similarly, for an input signaling a

change from the joy state to the disgust state, the baseline data showed a change of -0.1, while this test case produced a change of -0.17. A true comparison of the effect of sequential negative axis emotions on mood was limited due to the constraint of the mood model, which limits mood to a minimum value of 0.0, and based on the fact that mood was so universally low, a single transition into any negative axis emotion usually resulted in a mood value of 0.0.

Since psychological evidence indicates that individuals with a low Extroversion factor are expected to have a more negative mood, this test case supports Hypothesis Two.

In Test Case Ten, emotional response was uniformly negative. No input signaled a change into a positive emotion state. As such, intensity differed from that seen in the baseline data, but functioned as expected within the model. Similarly, mood was lowered as expected in the model, which differed from the baseline data.

Since psychological trials show that individuals with a lower Agreeableness factor show a higher incidence of negative emotion, the data from Test Case Ten clearly supports this prediction and therefore, this test case supports Hypothesis Two.

In Test Case Eleven, both emotion and intensity response were completely bottomed out. This also impacted mood, since the character never transitioned out of its initial state of joy, but the impact in mood response is as defined in the model. As expected from psychological data, a low Neurosis factor produces an emotionally flat character, and therefore Test Case Eleven supports Hypothesis Two.

Since every test case evaluated by method one support Hypothesis Two, and since no other test cases can be found to evaluate the method, Hypothesis Two is accepted.

Therefore, since both Hypotheses have been accepted, and since no counter examples were found, it is concluded that this project successfully created psychologically relevant implementations of synthetic personality, and synthetic personality dynamics.

CHAPTER FIVE: FUTURE WORK

The long term goal of this line of research is to create characters that are able to successfully pass the Turing test with regard to intentionality. The first step along these lines is to create and verify the models and is accomplished in this work. Following this verification, the cognitive models could be expanded to include more robust behavior and more expressive responses. To that end, the CBR implemented in the current work could be expanded to implement a more complex lexical matching scheme and a more expressive response generation scheme.

Lexical matching in the expanded model would follow the algorithm specified below:

1. For each user query, a query tree will be created. This tree will contain three sub-trees of queries, defined by:

   a. Starting with the full sentence, recursively replacing the last word in the sentence with an "anything" token (an asterisk). This will be called the "minimal" sub-tree.

   b. Starting with the full sentence, recursively replacing the first word in the sentence with an "anything" token. This will be called the "expanded" sub-tree.

   c. Prefixing each query in the minimal sub-tree with an "anything" token. This will be called the "maximal" sub-tree.

2. For each sub-tree, minimal, expanded, and maximal, the case library will be traversed, searching for a match.

3. The response will be selected based on the depth into each sub-tree, with the preference given to the shallowest tree.

4. In the case of a tie, the preference will be given to the minimal sub-tree, then the expanded sub-tree.

For example, if the user input is: "I want a piece of cake", the minimal sub-tree would be:

1. I want a piece of cake

2. I want a piece of *

3. I want a piece *

4. I want a *

5. I want *

6. I *

7. *

The expanded sub-tree would be:

1. I want a piece of cake

2. * want a piece of cake

3. * a piece of cake

4. * piece of cake

5. * of cake

6. * cake

7. *

The maximal sub-tree would be:

1. * I want a piece of cake

2. * I want a piece of *

3. * I want a piece *

4. * I want a *

5. * I want *

6. * I *

7. *

By searching for the most explicit match first, the system would be optimized towards finding high-fidelity matches. By setting the precedence as minimal-expanded-maximal, the author would ensure that same-subject phrases are matched before different-subject phrases, and that complete sentence matches would take priority over phrase-only matches. This algorithm implements lexical matching of sentences and would further allow the synthetic character to display lexical intentionality and automatic cognition. It also would better implement Staats' psychological behaviorism theory and would fill the role of the linguistic-cognitive repertoires for the current work.

To fulfill the 'new case adoption' requirement of the CBR algorithm, a system of analogical learning could be implemented in which novel patterns that do not have current matches in the case library would be added to the knowledge base. This learning system would be implemented lexically by allowing the system to match synonyms in the stimulus and create new cases by combining the new stimuli with the response of the synonymous case found. For example, if the case identified by the stimulus-response pair: "I want a piece of cake"-"I love cake!" exists in the case library, and a novel stimulus: "I desire a piece of cake" is presented without an explicit match in the database, a new case that contains: "I desire a piece of cake"-"I love cake!" will be added to the case library. Similarly, the pattern-response pair: "I want a piece of cake"-"I love cake!" could be transformed to fit the novel stimuli: "I want a piece of

92

pie", by substituting the word "pie" for the word "cake" in the original case. These synonym-based transformations could be accomplished through the use of the WordNet lexical database for the English language (WordNet, 2004).

The analogical learning approach specified in the previous paragraph would also partially fulfill the creative problem solving requirement of the cognitive model. The addition of a set of production rules that are capable of more robust pattern-response transformations, semantic context mapping, and understanding would also be required.

Once the models of personality and personality dynamics have been verified, next steps include testing the models with human users, with possible hypotheses being:

1. Participants will be able to correctly label the emotional state of synthetic characters based on pre-generated dialog.

2. Participants will be able to correctly label the general personality types of synthetic characters based on pre-generated dialog.

3. Participants will report a higher sense of engagement, interactivity, and realism for synthetic characters with the ability to communicate lexical and creative intentionality than for those characters without this ability based on real-time, constraint response conversation.

A survey could be designed to verify that the intention of the character is communicated. Questions would assess both positive and negative responses. That is, if the response is meant to communicate that the character is in a good mood and is highly extroverted, the user would be asked if he agrees that the character is friendly, and that he disagrees that the character is reserved or aloof. These types of questions could be answered with a Likart scale, in which "I don't know" is identified at a no-value answer. Additionally, the surveys would explicitly define

the terms being used to avoid confusion. The results of these surveys would identify whether the characters can convey intentionality appropriately while isolating negative experience factors that might be introduced by interacting with a computer.

Results from this study would identify whether the theories advanced by the author in terms of personality dynamics are effective. An additional study meant to identify whether these theories can be applied by a synthetic character in a conversation held in real time could follow. In this manner, successes and failures can be properly identified and the causal factors determined.

LIST OF REFERENCES

A.L.I.C.E. 2004. http://www.alicebot.org

Ajjanagadde, V. Incorporating Background Knowledge and Structured Explanada in Abductive Reasoning: A Framework. 1993. In *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 23, Number 3:. 650 – 664.

Ashcraft, M. 1998. Fundamentals of Cognition. Addison-Wesley Longman, Inc.

Badler, N., M. Costa and Widiger, L. Zhao, D. Chi. 2000. To Gesture or Not to Gesture: What is the Question? In *Proceedings of Computer Graphics International*: 3—9. Geneva, Switzerland.

Chen, L., K. Bechkoum, and G. Clapworthy. 2001. Equipping a Lifelike Animated Agent with a Mind. In *Intelligent Virtual Agents 2001*: 72—85. Madrid, Spain.

Chen, L., K. Bechkoum, G. Clapworthy. 2001. A Logical Approach to High-Level Agent Control. In *AGENTS'01*: 1 - 8,. Montreal, Quebec, Canada.

Clark, L.A., Vorhies, L., McEwen, J. 2002. Chapter 8. In Personality Disorders and the Five Factor Model: Second Edition, Ed. Paul T. Costa and Widigers& Thomas A. Widigier, American Psychology Association, Washington D.C., USA

Cloninger, S. 2004. Theories of Personality: Understanding Persons (fourth edition) Pearson Prentice-Hall.

Costa, P.T. Jr., Widiger, T. 2002. Chapter 1. In Personality Disorders and the Five Factor Model: Second Edition, Ed. Paul T. Costa and Widigers& Thomas A. Widigier, American Psychology Association, Washington D.C., USA

Douglas, Y., A. Hargadon. 2000. The pleasure principle: immersion, engagement, flow. In *Proceedings of the eleventh ACM on Hypertext and hypermedia*: 153—160. San Antonio, TX, USA

Ekman, P. 1992. An argument for basic emotions. In *Cognition and Emotion 6*: 169—2000.

El Nasir, M., T. Ioerger, J. Yen. 1999. PETEEI: A Pet with Evolving Emotional Intelligence. In *Autonomous Agents '99*: 9—16. Seattle, WA, USA

Emmons, R., E. Deiner. 1986. An interactional approach to the study of personality and emotion. In *Journal of Personality*, Vol. 84, Issue 2. 371-395.

Ferber, J. 1999. <u>Multi-Agent Systems, An Introduction to Distributed Artificial Intelligence</u>.
  Addison-Wesley.

Fox, R. 1999. A generic task strategy for solving routine decision making problems. In *Tools
  with Artificial Intelligence*:. 319 – 322. Chicago, IL, USA

Fujita, F., 2004. http://www.iusb.edu/~ffujita/Documents/big5.html

Gonzalez, A. 2002. Class lecture in EEL 6876: Advanced Topics in Artificial Intelligence.
  University of Central Florida.

Gratch, J., S. Marsella. 2001. Tears and Fears: Modeling Emotions and Emotional Behavior in
  Synthetic Agents. In *AGENTS '01*: 278—286. Montreal, Quebec, Canada.

Green, D., R. Cooper, R. Carston, A. Johnston, P. Howell, J. Morton, H. van de Koot, D.
  O'Neill, D. Shank and Abelmans, P. Burgess. 1996. <u>Cognitive Science: An Introduction</u>.
  Blackwell Publishers, Inc.

Gross, J., 2002. Chapter 21: Emotion and Emotion Regulation. In <u>Personality Disorders and
  the Five Factor Model: Second Edition</u>, Ed. Paul T. Costa and Widigers& Thomas A.
  Widigier, American Psychology Association, Washington D.C., USA

Hayes-Roth, B. 1997. Acting in Character. In *Creating Personalities for Synthetic Characters.*
  Eds. R Trappl and P. Petta. Springer.

Hibbard, B. 2002. Social Synthetic Characters. In *ACM SIGGRAPH Computer Graphics*, Vol.
  36, Issue 2. 5-7.

Hillerbrand, E. 1989. Cognitive Differences between Experts and Novices: Implications for
  Group Supervision. In *Journal of Counseling and Development*, Vol. 67, Issue 5. 293-297.

Hudlicka, E. 2002. This time with feeling: Integrated model of trait and state effects on
  cognition and behavior. In *Applied Artificial Intelligence*, Vol. 16, Issue 7-8. 611-642.

Hurlburt, R., M. Koch, C. Heavey. 2002. Descriptive experience sampling demonstrates the
  Connection of Thinking to Externally Observable Behavior. In *Cognitive Therapy and
  Research*, Vol. 26, Issue 1. 117-135.

Kille, M., K. Warwick. 2001. Queue-based Agent Communication Architecture for Intelligent,
  Human-like Computer Intefaces. In *Conference Publication 481: People in Control, An
  International Conference on Human Interfaces in Control Rooms, Cockpits, and Command
  Centers*: 206—212.

Klein, G. 1997. Chapter 27. In <u>Naturalistic Decision Making</u>, Ed. Caroline E. Zsambok & Gary
  Klein, Lawrence Erlbaum Associates, Inc.

Kokkonen, M., L. Pulkkinen. 2001. Examination of the paths between personality, current mood, its evaluation, and emotional regulation. In *European Journal of Personality*, Vol. 15, Issue 2. 83-104.

Kshirsagar, S., N. Magnenat-Thalmann. 2002. <u>A Multilayer Personality Model</u>. In *Symposium on Smart Graphics*: 107-115. Hawthorne, NY, USA

Luger, F., W. Stubblefield. 1993. <u>Artificial Intelligence: Structures and Strategies for Complex Problem Solving, Second Edition</u>. The Benjamin/Cummings Publishing Company, Inc.

McKeown, K. 1985. <u>Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text</u>. Cambridge University Press.

Missiaen, L. 1991. Localized Abductive Planning for Robot Assembly. In *IEEE International Conference on Robotics and Automation*: 605 – 610. Sacramento, CA, USA

O'Malley, M., C. Gillet. 1984. Exploring the relation between traits and emotions. In *Journal of Personality*, Vol. 52, Issue 3. 274-285.

Paiva, A., I. Machado, R. Prada. 2001. Heros, Villains, Magicians… Dramatis Personae in a Virtual Story Creation Environment. In *Proceedings of the 6th International Conference on Intelligent User Interfaces 2001*: 129--136. Santa Fe, New Mexico, USA.

Poggi, I. 2001. The Lexicon and the Alphabet of Gesture, Gaze, and Touch. In *Intelligent Virtual Agents 2001*: 235--236. Madrid, Spain.

Reilly, W.S., J. Bates. 1995. Natural Negotiation for Believable Agents. Tech. Report: CMU-CS-95-164. Carnegie Mellon.

Reeves, B., C. Nass. 1996. <u>The Media Equation: How People Treat Their Computers, Television, and New Media like Real People and Places</u>. Cambridge Press.

Rickel, J. 2001. Intelligent Virtual Agents for Education and Training: Opportunities and Challenges. In *Intelligent Virtual Agents 2001*: 15—22. Madrid, Spain.

Rousseau, D., B. Hayes-Roth. 1998. A Social-psychological Model for Synthetic Actors. In *Proceedings of the Second International Conference on Autonomous Agents*: 165—172. Minneapolis, MN, USA

Satoh, K., K. Yamamoto. 2002. Speculative Computation with Multi-Agent Belief Revision. In *AAMAS'02:* 897 – 904. Bologna, Italy.

Shank, R., R. Abelman. 1977. <u>Scripts, Plans, Goals, and Understanding</u>. Laurence Erlbaum Associates.

Solomon, R. 2002. Back to Basics: On the Very Idea of "Basic Emotions". In *Journal for the Theory of Social Behavior*, Vol. 32, Issue 2. 115-145.

Vel'asquez, Juan D. 1997. Modeling Emotions and Other Motivations in Synthetic Agents. In *AAAi'97:* 10-15.

Watts, F. 1997. Psychological and religious perspectives on emotion. In *Zygon: Journal of Religion and Science*, Vol. 32, Issue 2. 243-261.

WordNet. 2004. http://www.cogsci.princeton.edu/~wn/index.shtml

Yoon, S., B. Blumberg, and G. Schneider. 2000. Motivation driven learning for interactive synthetic characters. In Proceedings of the fourth international conference on Autonomous agents: 365—372. Barcelona, Spain.