University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

Spring 5-16-2011

Modeling Memes: A Memetic View of Affordance Learning

Benjamin D. Nye
University of Pennsylvania, benjamid@seas.upenn.edu

Follow this and additional works at: http://repository.upenn.edu/edissertations

Part of the Artificial Intelligence and Robotics Commons, Cognition and Perception Commons, Other Ecology and Evolutionary Biology Commons, Other Operations Research, Systems Engineering and Industrial Engineering Commons, Social Psychology Commons, and the Statistical Models Commons

Recommended Citation
Nye, Benjamin D., "Modeling Memes: A Memetic View of Affordance Learning" (2011). *Publicly Accessible Penn Dissertations*. 336.
http://repository.upenn.edu/edissertations/336

With all thanks to my esteemed committee, Dr. Silverman, Dr. Smith, Dr. Carley, and Dr. Bordogna. Also, great thanks to the University of Pennsylvania for all the opportunities to perform research at such a revered institution.

This paper is posted at ScholarlyCommons. http://repository.upenn.edu/edissertations/336
For more information, please contact libraryrepository@pobox.upenn.edu.

Modeling Memes: A Memetic View of Affordance Learning

Abstract

This research employed systems social science inquiry to build a synthesis model that would be useful for modeling meme evolution. First, a formal definition of memes was proposed that balanced both ontological adequacy and empirical observability. Based on this definition, a systems model for meme evolution was synthesized from Shannon Information Theory and elements of Bandura's Social Cognitive Learning Theory. Research in perception, social psychology, learning, and communication were incorporated to explain the cognitive and environmental processes guiding meme evolution. By extending the PMFServ cognitive architecture, socio-cognitive agents were created who could simulate social learning of Gibson affordances. The PMFServ agent based model was used to examine two scenarios: a simulation to test for potential memes inside the Stanford Prison Experiment and a simulation of pro-US and anti-US meme competition within the fictional Hamariyah Iraqi village. The Stanford Prison Experiment simulation was designed, calibrated, and tested using the original Stanford Prison Experiment archival data. This scenario was used to study potential memes within a real-life context. The Stanford Prison Experiment simulation was complemented by internal and external validity testing. The Hamariyah Iraqi village was used to analyze meme competition in a fictional village based upon US Marine Corps human terrain data. This simulation demonstrated how the implemented system can infer the personality traits and contextual factors that cause certain agents to adopt pro-US or anti-US memes, using Gaussian mixture clustering analysis and cross-cluster analysis. Finally, this research identified significant gaps in empirical science with respect to studying memes. These roadblocks and their potential solutions are explored in the conclusions of this work.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Electrical & Systems Engineering

First Advisor

Barry G. Silverman, Ph.D.

Keywords

Memes, Affordances, Agent Based Model, Cognitive Modeling, Social Systems, Artificial Intelligence

Subject Categories

Artificial Intelligence and Robotics | Cognition and Perception | Other Ecology and Evolutionary Biology | Other Operations Research, Systems Engineering and Industrial Engineering | Social Psychology | Statistical Models

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/336

Comments
With all thanks to my esteemed committee, Dr. Silverman, Dr. Smith, Dr. Carley, and Dr. Bordogna. Also, great thanks to the University of Pennsylvania for all the opportunities to perform research at such a revered institution.

This dissertation is available at ScholarlyCommons: http://repository.upenn.edu/edissertations/336

MODELING MEMES: A MEMETIC VIEW OF AFFORDANCE LEARNING

Benjamin D. Nye

A DISSERTATION
in
Electrical and Systems Engineering
Presented to the Faculties of the University of Pennsylvania
in
Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy
2011

Supervisor of Dissertation: Dr. Barry Silverman, Professor

Graduate Group Chairperson: Dr. Roch Guerin, Alfred Fitler Moore Professor

Dissertation Committee

Chair:	Dr. Tony Smith, University of Pennsylvania
Reader:	Dr. Joseph Bordogna, University of Pennsylvania
Reader:	Dr. Kathleen Carley, Carnegie Mellon University

Modeling Memes: A Memetic View of Affordance Learning
Copywrite © Benjamin Daniel Nye, 2011

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. To view a copy of this license, visit:
http://creativecommons.org/licenses/by-nc-nd/3.0/
or send a letter to: Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

Acknowledgments

My greatest thanks and regards to my advisor Dr. Barry Silverman, the members of the ACASA Lab past and present, and my family who has been fully behind me since the day I was born (and perhaps before). I would also like to thank my committee: Dr. Tony Smith, Dr. Joseph Bordogna, and Dr. Kathleen Carley. I am honored to have such approachable and distinguished members consulting me on this endeavor. Each of your insights has made me examine different perspectives that have improved not only this research, but will also guide me on my next steps.

Additionally, I would like to thank all the external scholars who have assisted me on this journey. For the data from the Stanford Prison Experiment, I received an unprecedented level of assistance. Dr. Zimbardo is one of the most approachable scholars I have ever contacted, extremely responsive and helpful. The members of the Archives of the History of American Psychology also graciously let me work on-site for many days as I collected data. I would like to give a special thanks to Rhonda Rinehart for being my point of contact and also to Dr. Baker for helping to arrange my access. Without such assistance, I could not have worked with such landmark data.

I would also like to thank my wife, Yuko, who has been patient with me through all the long days and nights toward the completion of this effort. I will always appreciate your love and support through this period. Your mere presence means the world to me.

Finally, I would like to thank God for giving me the resources and stability in life to pursue these endeavors, which I hope will someday make some small benefit to the world. I promise that this work is only the beginning of a long and dedicated effort toward improving our understanding and interaction with the world.

Modeling Memes:
A Memetic View of Affordance Learning

Benjamin D. Nye
Barry G. Silverman

Abstract

This research employed systems social science inquiry to build a synthesis model that would be useful for modeling meme evolution. First, a formal definition of memes was proposed that balances both ontological adequacy and empirical observability. Based on this definition, a systems model for meme evolution was synthesized from Shannon Information Theory and elements of Bandura's Social Cognitive Learning Theory. Research in perception, social psychology, learning, and communication were incorporated to explain the cognitive and environmental processes guiding meme evolution. By extending the PMFServ cognitive architecture, socio-cognitive agents were created who could simulate social learning of Gibson affordances. The PMFServ agent based model was used to examine two scenarios: a simulation to test for potential memes inside the Stanford Prison Experiment and a simulation of pro-US and anti-US meme competition within the fictional Hamariyah Iraqi village. The Stanford Prison Experiment simulation was designed, calibrated, and tested using the original Stanford Prison Experiment archival data. This scenario was used to study potential memes within a real-life context. The Stanford Prison Experiment simulation was complemented by internal and external validity testing. The Hamariyah Iraqi village was used to analyze meme competition in a fictional village based upon US Marine Corps human terrain data. This simulation demonstrated how the implemented system can infer the personality traits and contextual factors that cause certain agents to adopt pro-US or anti-US memes, using Gaussian mixture clustering analysis and cross-cluster analysis. Finally, this research identified significant gaps in empirical science with respect to studying memes. These roadblocks and their potential solutions are explored in the conclusions of this work.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

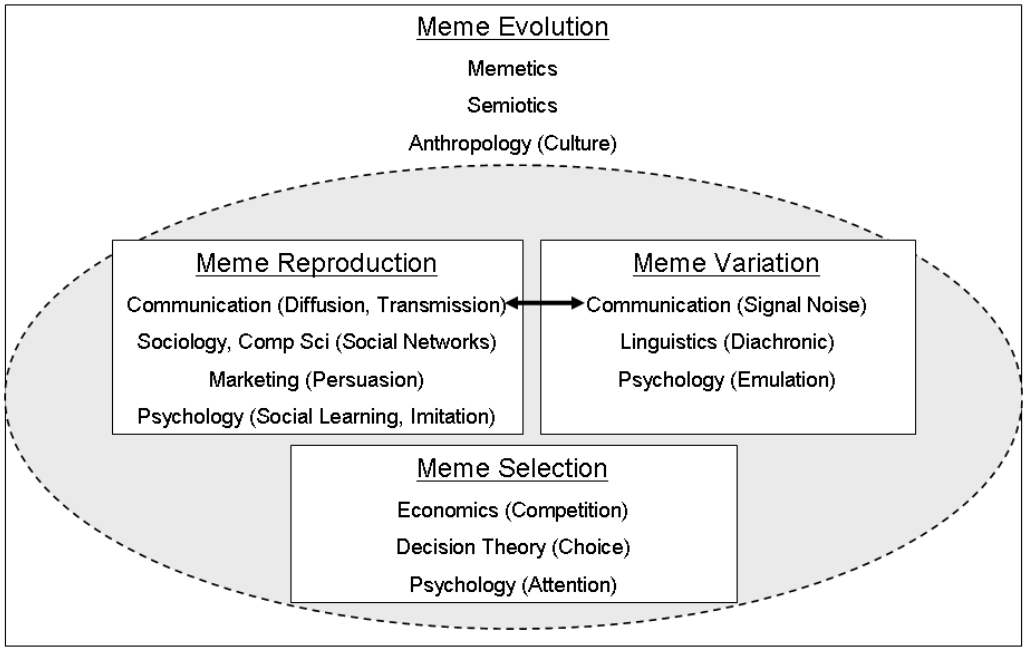
How does culture spread and change? Despite being a salient issue since the times of Plato and Confucius, the processes underlying cultural shifts remain intractable. Science cannot forecast cultural changes, nor can it consistently formulate the problem tractably (Koomey, 2002). Even identifying the data required to study cultural changes remains an open question. Memetics offers a potential solution, treating culture as an ecosystem of evolving ideas. In this framework, each meme is analogous to a "species" of cultural information that reproduces, mutates, and competes within the social system. As such, memes link individual behavior and cognition with the dynamics of culture. To study memes effectively, memetics requires rigorous definitions and interfacing with cognitive approaches (Castelfranchi, 2001). Systems theory provides a framework which makes functional study of memes possible.

This paper presents a synthesis of systems theories into a useful model for meme evolution. Shannon's Information Theory and Bandura's Social Learning Theory are central to this architecture (Bandura, 1986; Shannon, 1948). These theories provide complementary processes for examining the flow of information between and within individuals, respectively. Semiotics and evolutionary theory are examined for their insight into the workings of memes. With a rigorous description of memes in hand, experiments and theory from cognitive psychology, social psychology, and perception are used to introduce a semantic layer for understanding memes. Finally, this model was implemented as an agent based model and used to simulate real world situations.

1.1 Studying Memetics Using the Systems Social Science Paradigm

Memes are a relatively new topic in modern science, introduced by philosopher Richard Dawkins in the *The Selfish Gene* (Dawkins, 1976). The seminal works on memes continue to be produced by philosophers, such as in *Darwin's Dangerous Idea* by Dennett (1995) and *The Meme Machine* by Blackmore (1999). While this lineage has yielded interesting concepts and propositions, one criticism of *The Meme Machine* extends to much of memetics: "There isn't a lot on the workings of the memetic machinery" (Gabora, 1999). The study of memetics has opened questions almost exclusively, with even the rigorous definition of a meme remaining unclear (Blackmore, 1999, p. 52). While scholars studying memes agree that memes represent a pattern or behavior that is copied from one person to another, some scholars continue to debate the definition of memes with little consideration toward applying the concept to real world issues.

Figure 1.1: Fields Empirically Studying Memes



Meanwhile, a wealth of relevant empirical research has been waiting on the sidelines without being applied memes. Memes have considerable promise, but they are a truly interdisciplinary topic that requires a systems approach. Ongoing work in a variety of fields have implications for the dynamics of memes. Figure 1.1 outlines some of the key fields which are studying all or part of the meme evolution process. Just as these fields assist in understanding memes, memes can help to understand these fields. Unfortunately, social science data has primarily been analyzed within narrow domains, forming their own disconnected knowledge hierarchies (Silverman, 2004). The understanding provided by these is deep but narrow. Understanding memes requires extracting the best wisdom out of social science theories and using it as the knowledge base for memes.

To organize this information, a systems social science approach was employed (Silverman, 2010). This process works by synthesizing existing social science theories and narrowly scoped empirical specialties (silos) into a synthesis model that can applied to broader social issues. This approach complements reductive approaches to studying cognition and social systems. Traditionally, reductive analysis tries to reduce a system down to a set of parts and examine each part. The systems social science approach focuses on “synthesis” rather than focusing on “analysis.” Synthesis attempts understand the purpose of each part of the system and the inter-relationships between these parts. In this way, the systems social science approach has been used to integrate the results of reductive analysis and explore their implications on memes within social systems.

Figure 1.2: Systems Social Science Development Cycle. Adapted from Silverman (2006)

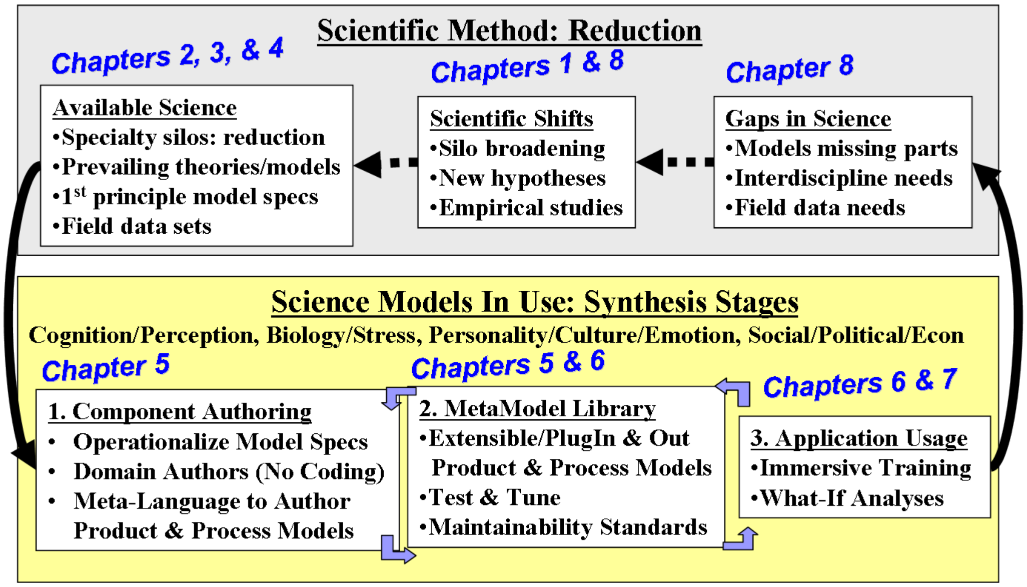
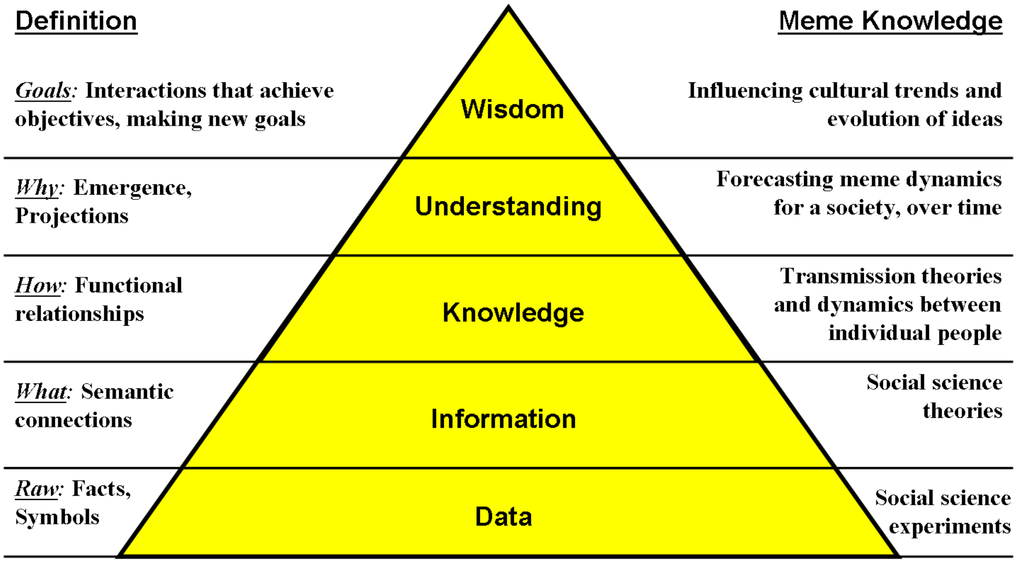

Scientific Method: Reduction
Chapters 2, 3, & 4
Chapters 1 & 8
Chapter 8
Available Science
•Specialty silos: reduction
•Prevailing theories/models
•1st principle model specs
•Field data sets
Scientific Shifts
•Silo broadening
•New hypotheses
•Empirical studies
Gaps in Science
•Models missing parts
•Interdiscipline needs
•Field data needs
Science Models In Use: Synthesis Stages
Cognition/Perception, Biology/Stress, Personality/Culture/Emotion, Social/Political/Econ
Chapter 5
Chapters 5 & 6
Chapters 6 & 7
1. Component Authoring
• Operationalize Model Specs
• Domain Authors (No Coding)
• Meta-Language to Author Product & Process Models
2. MetaModel Library
•Extensible/PlugIn & Out Product & Process Models
•Test & Tune
•Maintainability Standards
3. Application Usage
•Immersive Training
•What-If Analyses

Figure 1.2 shows the development cycle for how focused social science findings can be operationalized and synthesized into a common framework for social systems inquiry. This figure was adapted from Silverman (2006) and has been overlaid with the chapters that describe each step of the development cycle.

This research started with a stated goal of silo-broadening: expanding memetics to include key empirical research that explains the environmental and cognitive mechanisms that underlie memes (a scientific shift). This chapter describes a plan of research to broaden how memes are approached as a science. In the conclusion in Chapter 8, the paper returns to this stage to with new hypotheses and ideas for research on this topic.

To accomplish this goal, empirical research from the domains stated in Figure 1.1 was examined and applied to the problem (available science). Chapters 2, 3, and 4 each examine different theoretical and empirical concerns related to memes and how they can be studied as part of a social system. As part of this process, the theoretical findings were organized into a conceptual framework, which occurs as part of the transition between "available science" and "component authoring." Insight from these theories and findings was integrated into a systems model for memes, at the conceptual level.

Figure 1.3: Data to Wisdom: Toward Understanding Memes


Definition
Meme Knowledge
Goals: Interactions that achieve objectives, making new goals
Wisdom
Influencing cultural trends and evolution of ideas
Why: Emergence, Projections
Understanding
Forecasting meme dynamics for a society, over time
How: Functional relationships
Knowledge
Transmission theories and dynamics between individual people
What: Semantic connections
Information
Social science theories
Raw: Facts, Symbols
Data
Social science experiments

Synthesizing a model for memes harnesses insight from social science research. This is partly a knowledge management problem: information related to memes has not been used to understand memes. Russell Ackoff's (1989) knowledge hierarchy defines five levels: data, information, knowledge, understanding, and wisdom. Figure 1.3 shows the knowledge hierarchy with regard to memes. Little knowledge and minimal understanding of memes has been achieved, evident in the lack of practical applications for memes. A schism between information related to memes and meme knowledge would appear responsible for these limitations. The primary goal of this work is to bridge this gap, using the Shannon (1948) information model and Bandura (1986) social cognitive learning model as knowledge frameworks to aggregate information from social science theories.

Chapter 5 outlines how the systems model was used to create a computational implementation. This required operationalizing social science theories and models into component submodels (component authoring). These components were added as plug-ins to a larger modeling framework, which for this research was PMFServ (meta-model library). Based upon this expanded framework, individual scenarios were simulated and examined to study memes (application usage). Chapter 6 describes two scenario designs: a simulated Stanford Prison Experiment and an Iraqi village based on human terrain data. From applying the model to these scenarios, two types of insight were produced. The first insights are the direct findings from the simulations (what-if analyses). These findings are presented in the results section, Chapter 7. The second insights are the gaps in science which must be filled in order to expand the ability of the system to improve scientific inquiry (gaps in science). These gaps highlight avenues for new empirical studies and hypotheses to test, in order to expand knowledge about the mechanisms that influence meme evolution. The conclusions in Chapter 8 outline the gaps discovered in social science that impede the study of memes, the new hypotheses that could be tested, and some possible social science experiments that could greatly improve understanding and modeling meme evolution.

1.2 Objectives

Five core questions are approached in this paper:

1. Definition: What is a meme?
2. Systems Model: What synthesis of theories usefully explains meme evolution?
3. Measurement: How can memes be identified and measured empirically?
4. Implementation: Can the systems model be operationalized into a computational model?
5. Usefulness: Can the model be used to study real-world scenarios?

Resolving these questions requires a full community's labor, but the goals of this research are more modest: workable answers to build a useful architecture for memes. These key questions have not been adequately addressed for memetics, which has been limited by its disconnection from the extensive body of related empirical work. In addressing these questions, equal focus has been placed on theoretical soundness and applicability to empirical situations. A full chapter has been devoted to each of these fundamental questions, gradually working from theoretical concerns to practical applications. Assumptions and simplifications have been necessary, with all possible effort made to state where and why they were made. The result is not the final answer for modeling memes, but it does present a useful architecture for examining memes.

1.2.1 Definition: Formally Defining Memes

The first question addressed is the definition of a meme, without which any further discussion would be baseless. This question depends upon further questions: what are the elements involved in the meme system, what are their key relationships, and what are necessary and sufficient conditions for a meme to exist in this system? Chapter 2 presents a functional, empirically-approachable definition to a meme. This contribution begins by presenting scholarly perspectives on memes, considering the disparity between different definitions of a meme (Section 2.1). From this discussion, the concept of a meme is defined in terms of its semantic information (Section 2.2). Finally, this chapter presents a formal definition for memes that describes memes in terms of their ability to recursively reproduce within a society (Sections 2.3 and 2.4). This definition is used as a foundation for exploring issues of detecting, modeling, and measuring memes.

1.2.2 Systems Model: Synthesis of Theories to Explains Meme Evolution

Building up from the definition in Chapter 2, a systems approach for modeling meme evolution is presented. Chapter 3 describes a synthesis of the Shannon (1948) information model with Bandura's (1986) social learning model, complemented by additional models of human cognition and social psychology. Within this framework, meme evolution is emergent from each person's interaction with the environment as described in Section 3.1. Section 3.2 fits these individual interactions into framework formed by synthesizing Bandura's (1986) observational learning process with the Shannon (1948) information transmission model. This synthetic model represents both the physical and cognitive processing of memes. Using these aspects of the Social Cognitive Learning Theory (Bandura, 1986) and Information Theory (Shannon, 1948), this chapter describes the ecosystem for memes.

Within this framework, memes are treated as semantic information. By treating memes as information, the Shannon (1948) information model provides an architecture for analysis. The Shannon model has a simple design which allows for analysis of message transmission, which is one mechanism involved in memes. The Bandura (1986) model of social learning provides an overlapping and complementary level of analysis, focusing on how humans process of socially transmitted information. These models provide good starting points but still have limited explanatory power without more fine-grained and implementable mechanisms. To this end, cognitive theories of attention, motivation, and social psychology were integrated into the framework.

The remainder of Chapter 3 interprets how these cognitive theories work as mechanisms that guide meme evolution. The mechanisms behind the evolution of memes are the primary focus of Sections 3.3 and 3.4. The key questions are: Can meme behavior be explained by building up from existing theories and science? If so, what synthesis of theories are sufficient to explain meme dynamics? Through an extensive literature review, theories have been identified and synthesized to help explain meme dynamics. Section 3.3 describes how information theory can be used to explain meme variation and competition between memes. Section 3.4 introduces the cognitive "memetic machinery" of social learning theory, describing how empirically derived cognitive mechanisms influence meme reproduction, variation, and competition.

This synthesis model for memes is intended to be a useful model for studying meme evolution. A systems model is useful if it is complete, holistic, and workable. To be complete, the model must be able to meaningfully represent all three mechanisms for meme evolution: reproduction, variation, and selection pressure. To be holistic, the synthesis of Social Learning Theory (Bandura, 1986) and Information Theory (Shannon, 1948) must be essential for modeling meme evolution as opposed to using either theory in isolation. This integration must provide different insight than the set of disconnected theories. The conclusion of Chapter 3, Section 3.5, presents a fleshed-out conceptual model for memes that consolidates the mechanisms that affect meme evolution. Chapter 3 demonstrates that the model is complete and holistic. However, to show that it is workable the model must be applicable to real-world problems.

1.2.3 Measurement: Identifying and Measuring Memes Empirically

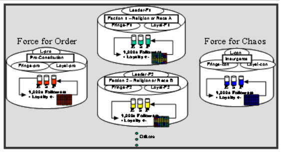
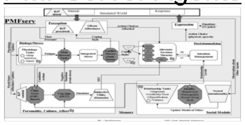
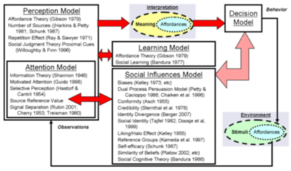
Chapter 4 shows that the definition and synthesis model for memes can be used to define measurements of memes. These measurements must be strict enough to be falsifiable, otherwise memetics do not form a well constrained field. The measurements presented in this section focus on proving that a meme exists and present methods for measuring meme reproduction. While focusing on general concepts, as opposed to the nitty-gritty of a specific experimental design, this chapter provides an outline of approaches to empirically studying different kinds of memes. Section 4.2 in this chapter also introduces a special type of meme: socially learned affordances (action possibilities). J. J. Gibson's (1979) affordance theory of perception posits that organisms perceive their environment in terms of the actions it affords. Action affordance learning provides a type of well structured meme that corresponds directly with a behavioral expression. This makes socially learned affordances particularly amenable to measurement and testing models against empirical data.

1.2.4 Implementation: Realizing the Memes Computationally

Having outlined the fundamental concepts for studying memes, the following chapters focus on transitioning this theory into a workable implementation. Chapter

5 outlines a computational implementation of the conceptual model defined in Section 3.5. The realization of the conceptual model is an agent based model, consisting of multiple autonomous and interacting agents. This computational model is designed to simulate social learning of affordances within an agent based simulation, due to their ability to be measured empirically (as detailed in Section 4.2). This particular computational implementation is designed to focus on the dynamics of meme reproduction rather than meme mutation. While mutation is also important, this simplification allows for correspondence tests against data sets collected to examine diffusion of innovation.

Figure 1.4: Levels of Analysis for Memes

Level of Analysis	***Modeling Tool***	***Purpose***
Socio-Cultural	FactionSim Simulation 	-**Study the spread of memes** in a society of inter-related agents, belonging to various groups. -**Examine emergent behavior** of social psychology models.
Individual	PMFServer Agents 	-**Connect new cognitive models** to build a model integrating the new models with existing models of emotion, stress, and decision.
Cognitive Models	Python Code 	-New **perception model** capable of social learning of actions. -**New descriptive models** of: - Attention - Social influences - Social attitudes

The computational model is implemented as an extension of the PMFServ agent based architecture, as described in Section 5.1. The PMFServ paradigm is a model of models approach: each simulation consists of groups, each group consists of agents, and each agent consists of a combination of cognitive models based on social science literature (Silverman, 2010). As shown in Figure 1.4, these three levels of analysis correspond to information, knowledge, and understanding.

New computational cognitive components were implemented, formalizing the information captured by social science theories. These components are each described in Section 5.2. These components represent the reusable plug-ins described in the "Meta-Model Library" (Figure 1.2). These components represent specific theories and empirical relationships that support modeling for memes. Attention, social influence, and learning models were implemented and connected with existing cognitive models to make agents capable of learning affordances socially. The relationships between models are the knowledge layer for memes, giving the dynamics for meme processing by a single agent. Decisions and behavior are emergent from this system of cognitive models. This emergent behavior feeds into the society, allowing learning and imitation that reproduces the meme. Reproduction, diffusion, and immunity to a meme are emergent properties of the larger social system. By simulating real world scenarios and examining these emergent properties, this implementation has helped increase understanding of how a meme's dynamics relate to its society.

1.2.5 Usefulness: Applying the Model to Study Real-World Scenarios

In order to demonstrate that the systems model for memes is useful, the implemented computational model was applied to two scenarios with real-world relevance: an examination of potential memes in the Stanford Prison Experiment and a simulation of competing memes in a fictional Iraqi village. The process of selecting and creating these scenarios is explained in Chapter 6. A simulation of the Stanford Prison Experiment was designed, attempting to examine different hypotheses for the source of guards' use of "the hole" on prisoners and prisoners' use of resistance against the guards. Empirical data was used to train and validate the computational model for the Stanford Prison Experiment simulation, which was collected from the Archives of the History of American Psychology. The design of the Stanford Prison Experiment scenario is described in Section 6.1.

This Stanford Prison Experiment simulation was trained over a limited period of the simulation length, with the remainder used for external validity testing. The hypotheses were tested by examining whether learning of actions as memes improved the ability of the model to predict the order of when those actions would be expressed. This analysis used a new metric for sequence correspondence, which is a normalized variant of an inversion count. This new metric is explained briefly in Section 7.2.4 and in more detail in Appendix I.

The Hamariyah Iraqi village scenario was intended to model competition between memes. This scenario design is described in Section 6.2. The Hamariyah scenario tests the capability of the model to provide wisdom, a legitimate insight into a situation beyond that provided by the base model. Meme awareness and meme expression were examined within the village, in an attempt to find individual and contextual characteristics affecting meme dynamics. This scenario models a meme for giving information to US forces about insurgent activity and a competing meme for volunteering to plant an IED for insurgents. In this respect, the model is used to examine which agents gravitate toward learning and expressing each meme.

Each scenario is simulated across repeated runs in an experiment framework, which allows simulation and data analysis over a distribution of possible initial states and simulation paths (Nye, Roddy, Bharathy, & Silverman, 2007). These simulations yield insight into the workability and usefulness of both the realized and the conceptual model for simulation of memes. If the implemented model successfully provides insight into examining an empirical scenario and can be used to examine meme competition in a meaningful manner, the model for memes can be considered workable.

By this yardstick, the results of these simulations presented in Chapter 7 demonstrate that the systems model for memes is workable. Section 7.3 of the results demonstrates the internal validity of the model, showing that the implemented combination of models works appropriately in a testbed condition in Section 7.3.1. It then demonstrates and explains how different contexts can affect the apparent influence of different factors on meme transmission, in Section 7.3.2 and the following discussion. This internal validity testing shows interesting relationships which have empirical implications on their own merit.

The Stanford Prison Experiment scenario is examined for its external validity in the results Section 7.4.1. This scenario is subjected to a number of validity metrics chosen before simulating the experiment, to validate that the simulation represents important dynamics discovered in the empirical data from the experiment. The key metric applied to this data is one that examines the order that agents first perform certain actions, also known as the order of first expression. As explained in Chapter 4, the order of first expression can provide information about meme transmission if learning the meme is a limiting factor on expressing it. As such, a sequence ordering metric was used to determine if simulating certain actions as memes better represented the order of first expression.

Both the Stanford Prison Experiment simulation and the Hamariyah Iraqi village were also examined using traditional diffusion of innovation metrics, in sections 7.4.2 and 7.5.1 for the Stanford Prison Experiment scenario and Hamariyah scenario respectively. This analysis demonstrates that the implemented model can be used to perform diffusion of innovation experiments and provides insight into the workings of memes in these simulations. Additionally, both scenarios were examined to determine the meme adoption dynamics: which agents learned and expressed certain memes more quickly. This is an important analysis, since it provides insight that other models for memes lack: a solid representation of the situational and individual differences that allow memes to spread successfully through some individuals but not through others. Section 7.4.2 explores these differences by examining individual agents.

Finally, section 7.5.2 moves beyond the basic meme adoption dynamics and associates these differences with situational and personality factors that cause an agent to adopt certain memes quickly as opposed to avoiding them entirely. These traits are extracted from the simulation data using statistical techniques. These results demonstrate the ability of the implemented model to examine questions of meme competition, at the level of inferring what kinds of agents are drawn to certain memes. The results of this section are promising and show that models following this design may be able to successfully infer the individual and contextual differences that lead people to imitate certain behaviors rather than others.

1.3 Summary

By using the systems social science methodology and approaching memes in a structured way, this research presents a path forward for memes- a promising theoretical concept that has suffered due to its lack of integration with empirical work (Silverman, 2010). This thesis presents a formal definition for memes, a systems model for studying memes, approaches for measuring memes empirically, a computational implementation of the systems model for memes, and two simulated scenarios which highlight different capabilities of the computational model: comparison against empirical situations and analyzing meme competition. Each of these components represents a significant contribution to the discourse on memes and also explores questions of more general scientific importance in the process. The overall contribution is to demonstrate that the systems model for memes is complete, holistic, and workable for modeling meme evolution. This shows that the systems model for memes is useful, a scientifically meaningful approach for studying memes. However, there can be no meaningful approach to studying memes unless memes are explicitly defined. In the following chapter, a formalization for memes is presented to address this issue.

Chapter 2

Defining Memes

Before embarking on a thorough analysis of memes, they must be verified to be worthy of study. What are memes? Why are memes important? Memes provide an algorithmic mechanism for the spread and persistence of behaviors, language, and ideas within a population. Memes evolve and undergo natural selection, the same processes underlying gene survival (Dennett, 1995).

The evolution of memes is explained at length in *Darwin's Dangerous Idea* (Dennett, 1995), so only a short summary will be presented. Evolution requires three processes: inheritance, variation, and selection pressure (Darwin, 1902). The inheritance of cultural information is incontrovertible, forming the basis of the socio-cultural learning model and the ratchet model of culture (Bandura, 1986; Tomasello, 1999). Variation of information occurs at many stages, due to noise in transmission and individual differences when interpreting information. Selection pressure results from limited capability to process information, limited time to express memes, and limited motivation to express memes. All three of Darwin's conditions are observed, indicating an evolutionary process guiding cultural information.

Memes connect individual behavior and psychology with their emergent effects at the societal level. Understanding evolution and reproduction of memes would be a breakthrough for social science analysis. Public policy initiatives could be promoted with greater effectiveness by better targeting tipping-point demographics. At-risk demographics for copycat crimes such as school shootings or suicides could be identified, given a source incident. Additionally, punctuated equilibria for belief change could be forecasted- the birth of social norms. The ability to anticipate and reliably influence norms would revolutionize social science.

2.1 The Debate on Meme Definition

The precise definition of a meme remains a contested ontological question. Memes have been proposed as a philosophical lens, a scientific discipline, and at the center of a theory of mind (Dawkins, 1976; Heylighen, 1998; Blackmore, 1999). Heylighen (1998) treats the matter most similarly, examining how to quantitatively test for meme existence. This discussion treats memetics, the study of memes, as a discipline of the social sciences focusing on the evolution of cultural information. The Tomasello (1999) ratchet model of cultural change expresses the most similar view of culture, situating it in both the individuals minds and the artifacts of a society. Under this definition, culture is posited to change through an evolutionary process (not just "as if" such a process occurred) but is not claimed as the fundamental process for self or identity.

A further point of confusion is the very meaning of a meme. Dawkins' seminal definition established a meme as a "unit of cultural information," the internalist perspective (Dawkins, 1976). Adaptations to this definition make claims that such information must be able to be copied and recalled within the brain (Aunger, 2002). The externalist perspective frames memes in terms of their physical manifestations, such as behavior, messages, and signs. The systems perspective must adopt a semiotic view: that internal and external parts cannot be disentangled. Disconnecting the physical expressions from the cognitive information fundamentally breaks the meme replication process. This statement is controversial, but necessary- the evolution of a meme depends both on its physical manifestations and its cognitive interpretation.

The discussion from memes has also hit snags over the mental representations of memes within the brain, such as encoding techniques and neural localization. No stance will be taken over the internal representation of a meme. Studies of learning and memory in psychology have successfully expanded understanding of the cognition and behavior for over a century, despite only mapping neurons to responses for aplysia and squids (Marder, Abbott, Turrigiano, Liu, & Golowasch, 1996; Hodgkin & Huxley, 1952).

The problem of memory and mental representations would appear to be part of the general study of learning and memory- interesting in its own right, but not crucial to memetics. As an analogy, an undefragmented hard disk for a computer will regularly split up large files into many different sectors. If spatial contiguity is unnecessary for a file on a computer, why would it be necessary in a brain? It has also been argued that the localization approach may be one of opportunity, driven by the ability to measure the brain without understanding of the meaning of measurement (Uttal, 2001). If memes could be localized, very interesting experiments could be conducted. However, market researchers regularly measure awareness and attitudes toward ideas and products without resorting to neurological measures. Sidestepping the physical representation of a meme allows greater focus on the information characteristics and function of a meme.

2.2 Memes as Information

Memes are a special type of information. The term *information* must be used carefully, as memes involve multiple types of information. A meme must have a physical transmission (syntax) as well as a cognitive interpretation (semantic information). Peripheral and contextual information can overlay additional semantics. Dual process models of persuasion highlight the importance of these context cues, which can augment or entirely override the semantic content of a transmission (Petty & Cacioppo, 1986).

The core information of a meme is its semantic information. When semantic information changes, the meme has mutated or a new meme has been created. A meme reproduces when semantic information is replicated from one agent to another. Physical and contextual information accompany the meaning of a copied meme, but are generally absent when a meme is re-expressed. Copying a meme loses such information. For example, few people repeat jokes they heard in movies by repeating them in taped re-enactment of that movie. If the context becomes part of the meme, a variant meme has been created.

Though each new transmission has a new context, this context is very important for interpreting the physical transmission. Linguistic study addresses such issues of contextual understanding. The written and spoken versions of a word hold the same meaning despite differences in physical transmission. Conversely, identical physical transmissions change semantic meaning based on context (Gerot & Wignell, 1994). The word "embarazado" means "embarrassed" in Portuguese, but "pregnant" in Spanish. For a bilingual speaker, the surrounding words establish the meaning. The relationship between the receiver, syntax, and context will determine the received semantics- including memes (which are a subset of semantic information).

This relationship raises the issue of the connection between memes and the concept of signs in semiotics. Signs are a very general concept, defined generally as "A sign ... is something which stands to somebody for something in some respect or capacity" (Peirce, 1931, Vol 2, p.228). One opinion in the semiotics community views memes as "degenerate signs" because memes "copy" from one person to another rather than "translate" (Kull, 2000). While this has been a common approach to memes, there is no theoretical reason why memes should not need to be translated during the transmission process. Instead, memes can be viewed as a special subset of signs. Peirce (1931) states that an essential characteristic of signs is their potential to be interpreted, since a sign "would lose the character which renders it a sign if there were no interpretant." In other words, a sign must carry some meaning to a receiver in order to be a sign. A meme must also satisfy this characteristic or else it could never reproduce.

However, an interpretant (received meaning) must observe additional necessary conditions for its sign to be considered a meme. The information must be reproducible from one agent to another, the reproduction process must be a result of behavioral patterns, and some reproduced versions must remain reproducible. These conditions establish memes as a recursive case of observational learning (Bandura, 1977). The behavior or message specified by meme must be socially learned and capable of reproduction with fidelity.

2.3 Memes as Operators

A meme can be defined by its functional ability to sustainably reproduce within a society through social learning. This is similar to how computer viruses are defined as a subset of all possible combinations of code strings. In this view, a meme's semantic information contains a function definition. This approach to definition has precedent in biology's definition of life, in which reproduction is a necessary process (Koshland Jr., 2002). Wilkins (1998) expresses this condition for memes in his article *What's in a Meme?*

The necessary condition of recursive reproducibility can be expressed explicitly. The following symbols will be defined:

S - The society of agents under analysis.

Env - The environment that an agent inhabits.

Ω - All possible environments.

X_a - All information stored by an agent a.

x_a - Semantic information for a meme, as understood by agent a. $x_a \in X_a$.

$B_a(Env, X_a)$ - Behavior function of agent a. Alters and returns Env, returning it as a changed environment Env^*.

$B_a(Env, x_a)$ - Behavior function of an agent a when expressing x_a.

$P_a(Env, X_a)$ - Perception function of agent a. Alters and returns X_a, returning it as a changed set of stored information X_a^*.

The term "agent" is used to define members of a society, as it connects this definition to agent based simulation techniques explored later. At this stage, interchanging humans with agents must be discouraged, as significant evidence exists for cultural learning by apes, dolphins, and songbirds (Zentall, 2007).

The reproduction process of a meme is defined in Eqn. 2.1. A single reproduction requires three steps, each occurring over some span of time. First, the agent's relationship to the environment must activate some behavioral expression of the meme, changing their patterns of activity. Second, this behavior

Meme Reproduction Process

$$
\begin{array}{rl}
\textbf{Assume} & a_1, a_2 \in S \text{ s.t. } a_1 \neq a_2 \quad \& x_{a_1} \in X_{a_1} \quad \& \\
& \nexists\,(x_{a_2} \in X_{a_2}) \text{ s.t. } x_{a_2} \approx x_{a_1} \\
& \\
\textbf{Define} & R(a_1, a_2, Env): \\
& \quad B_{a_1}(Env, x) \in B_{a_1}(Env, X_{a_1}) \\
& \quad Env^* = B_{a_1}(Env, x) \\
& \quad X^*_{a_2} = P_a(Env^*, X_{a_2}) \\
& \quad \text{where } \exists (x_{a_2} \in X^*_{a_2}) \text{ s.t.}: x_{a_2} \approx x_{a_1}
\end{array}
\tag{2.1}
$$

Meme Reproduction Non-Triviality Condition

$$
\begin{array}{l}
P\{\exists x_{a_2} \in X^*_{a_2} : x_{a_2} \approx x_{a_1} \| B_{a_1}(Env, x) \in B_{a_1}(Env, X_{a_1})\} \gg \\
P\{\exists x_{a_2} \in X^*_{a_2} : x_{a_2} \approx x_{a_1} \| B_{a_1}(Env, x) \notin B_{a_1}(Env, X_{a_1})\}
\end{array}
\tag{2.2}
$$

must alter the environment in some observable way. Finally, a second agent must perceive this environment either during or after the behavior and learn new information similar to the meme. The question of what makes memes similar will be discussed further in Section 4.1 and in further depth in Appendix A. For reproduction (and the meme itself) to be non-trivial then Eqn. 2.2 must hold. Eqn. 2.2 formally states that the probability of learning the semantics would be much lower without observing the expression of the meme. This means that reproduction involves a transmission of semantic, meaningful information as opposed to a coincidental spontaneous learning event that would have been likely without such an observation.

From a certain standpoint, this definition of meme reproduction may seem overly general: any semantic information could be a meme, depending on the population and the environmental context. Indeed, if one makes no assumptions about the society or the environment- any information could be a meme. As a result, any meaningful study of memes must be tied to the society and environment. As the the environment and societal context is constrained, the set of possible memes becomes respectively constrained. As such, this definition of memes uses the society and environmental contexts almost like parameters- one cannot define memes without defining these two.

The society is important on many levels, depending on the level of specification given. The first obvious constraint is the species involved. For instance, humans appear capable of spreading a much different and wider array of memes than songbirds. Within a species, there are semantic requisites to learning certain information. For example, one cannot learn that an apple is on a dog without also learning about the concept of an apple and a dog. This may be thought of as a zone of proximal development for learning a meme, where knowing necessary concepts allows learning the meaningful information of the meme (Vygotsky, 1980). Through more detailed specification of the society, the set of possible memes can be constrained significantly. The environment, of which the communication medium is a part, also defines what memes can theoretically reproduce. In an extremely noisy environment, with no verbal communication possible, there would be significant constraints on the memes that could exist.

Even after maximally constraining the possible semantic information that can be reproduced, humans still have a vast amount of possible semantic information that can be communicated and reproduced. This definition may seem too general, bordering on a "theory of everything." Indeed, the reproduction process is general with respect to any semantic information. The special quality that differentiates memes from the larger body of semantic information is its ability to reproduce recursively- a significant additional constraint.

2.4 Recursive Reproduction of Memes

With reproduction defined, properties important to reproduction can be stated for the meme's viability. A necessary condition for semantic information to comprise a meme is the ability to reproduce recursively, stated formally in Eqn. 2.1. Information lacking such capability would be sterile and could not be called a meme. Note that a_1 and a_3 do not necessarily have to be different agents. The meme could pass back and forth, as when old friends relay an inside joke that they're prone to forgetting.

Figure 2.1: Recursive Reproduction Condition

$$\begin{gathered} Env_1, Env_2 \in \Omega \quad \& \quad \exists(a_1, a_2, a_3) \in S \quad \& \quad a_1 \neq a_2 \quad \& \quad a_2 \neq a_3 \\ \text{s.t. :} R(a_1, a_2, Env_1) \Rightarrow R(a_2, a_3, Env_2) \end{gathered} \tag{2.3}$$

The impact of a meme presented to an agent that has not forgotten their variant of the same meme does not have a clear effect. Any stimuli presented many times in sequence could create disconnected sets of information, a single updated set of associations, or a dynamic variation of connectivity. Such relationships are specific to an agent's cognitive processing; these are addressed as learning in Section 3.4.2. For simplicity, all variants similar enough to be labeled the same meme will be treated as a single meme. This assumption means that an agent cannot reproduce a meme to a carrier of the same meme but can update its meaning.

Recursive reproduction separates memes from other types of knowledge. The ability to successfully reproduce enables inheritance. This definition is ontologically complete: semantic information is a meme within a society and environment if and only if it can recursively reproduce in that society and environment. This may still overly general, but further restrictions on defining memes tend to be arbitrary.

For example, some alternative definitions further restrict the definition based upon how well the information reproduces or its ability to "motivate" reproduction Finkelstein (2008). This definition of reproduction places the cart before the horse- how can one measure how well a meme reproduces if one disregards inefficient memes? All memes would be, by definition, good at reproducing. Such restrictions on the definition appear counter-productive. Just like organisms, some species will reproduce better than others or be better suited for particular environments. This definition captures the theoretical ability of a meme to recursively reproduce, without coupling it to the motivational influence to reproduce or success in reproduction. These characteristics seem better suited to a continuum approach. Motivation to reproduce a meme is not binary, nor does it have to be consistent across members in a society. For this reason, these factors were kept separate from the definition of memes.

As a result, this definition of a meme provides no insight into what allows a meme to proliferate. By decoupling the definition from factors such as motivation and competition, these operationalizations serve to define what *could* spread as a meme rather than what *will* spread as a meme. Meaningful memetic study requires an understanding of the relationship between memes and their societal environment. The flow of information through a social system must be understood. In the following sections, a model is proposed that integrates the effects of cognition and the environment on meme transmission.

Chapter 3

The Meme Process: A Systems Model for Memes

A meme relies upon semantic information to propagate, meaning there can be no examination of memes that is disconnected from a population. Memes may be said to be cultural information not only because they augment culture but because their entire existence and meaning is predicated on the existence of a society where they may spread.[*] Members of a society provide an environment for memes to form an ecosystem.

Memes exist as part of a cultural system. In this context, a culture will be considered any collection of communicating individuals, their memes, and their semiotic signs. These three elements are interconnected and irreducible, as defined by Ackoff's definition of systems (Ackoff, 1971). Signs represent the communication of meaning through the physical environment. Memes and signs are meaningless without a society to interpret them. Similarly, a society incapable of both receiving and transmitting information between members could hardly be considered to communicate.

An analogy for such entanglement would be machine code, which has meaning only in relation to the chipsets of computers. Any random string of bits can in theory be machine code, given the proper instruction set. This analogy exposes the weakness of memes when no assumptions are made about the population's culture and cognitive processes. The strength of memetics can only come when a rich understanding of the actors is attained, either individually or as a societal distribution. Success in such an inquiry requires incorporating fundamental processes from psychology, behavioral economics, and other empirically active fields. In the words of Castelfranchi (2001), "Memetics needs cognitive modeling."

[*] After a talk on the UPenn campus, I managed to speak with Daniel Dennett and confirmed this point with him. A meme can exist only in relation to a population.

To embed cognitive models appropriately, memes must be expressed explicitly and rigorously at the systems level.

3.1 Interaction with the Environment

Agents and their environment form a closed system. Agents will be assumed to represent a boundary between the cognitive and physical domains, where agents have cognition and the environment does not. This approach does not represent a stance on the mind-body problem, but is taken in order to harness existing research on cognition. In real life, agents process information physically, but behavioral science traditionally examines cognition as an emergent process above this layer. To harness these findings, learning and mental processes will be treated cognitively.

Discussing the cognition of an agent requires a working definition of how an agent interfaces with their environment. Perception theory and control theory provide insight into the stimulus-response aspects of agent behavior. Perceptual theory posits that perception translates between the physical and cognitive worlds (James, 1890). The physical world cannot be fully perceived due to physical, cognitive, and even motivational limitations (Matthews & Wells, 1999).

Figure 3.1: Agent Perception Process

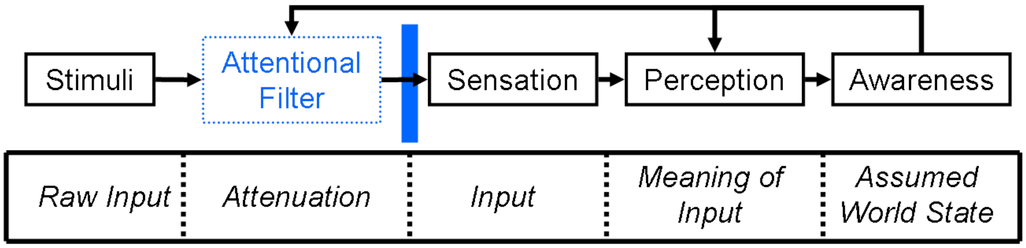
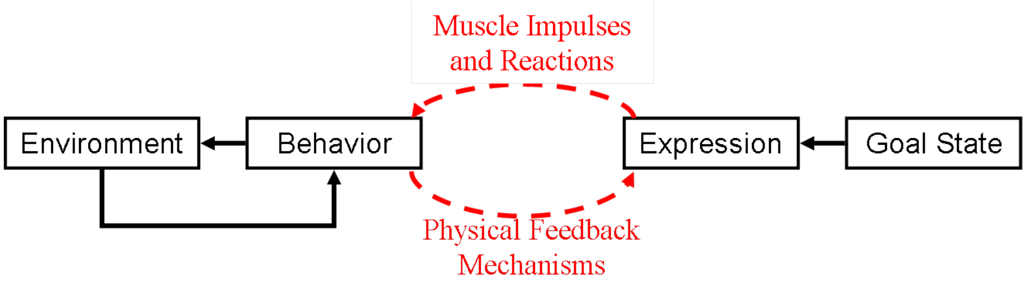


Figure 3.1 displays a diagram for the stages of perception. Stimuli represent the parts of the world that an agent can physically detect, given their sensory organs. Sensation represents what an agent detects from their sensory organs, that should convey some information about the stimuli. Attentional processes determine the set of stimuli that form sensations. Sensations must be interpreted to form perception, based upon current sensations and stored experiences. For an organism that learns, new perceptions will be stored as experiences and alter later perceptions.

The semantic information of a learned meme must be learned through perception. As a subset of learning, memes affect behavior and alter perception and interpretation of past events. Nothing unique about memes causes these changes, but the effects of existing memes on later memes should be noted. In the general case, meme information interacts path-dependently.

Figure 3.2: Behavioral Control Diagram

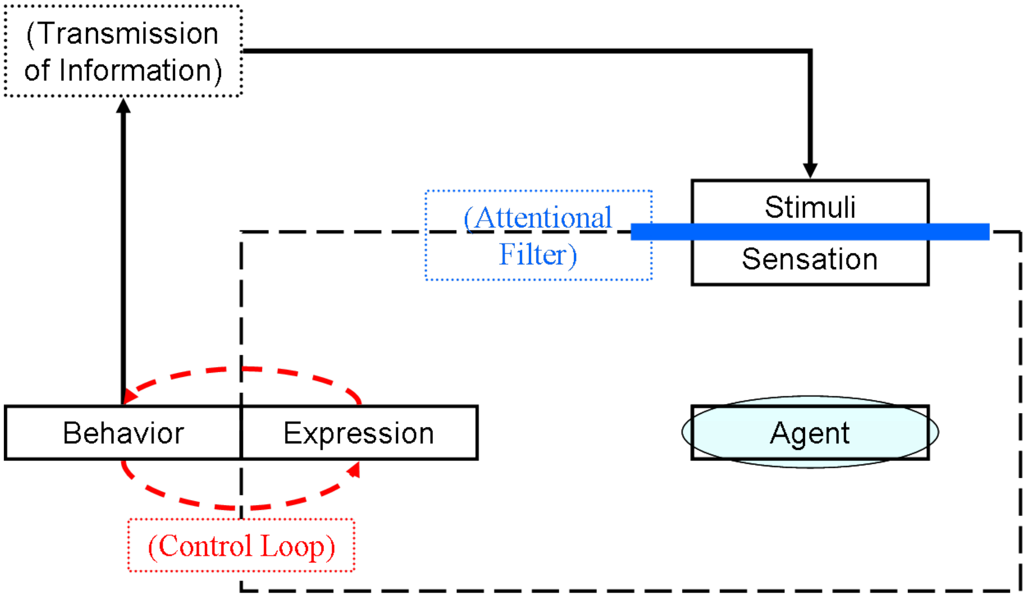


Behavior can be considered as a translation of cognitive impulses to physical responses. Behavioral expression is necessary for meme transmission. Control theory for systems applies to behavioral expression (Miall & Wolpert, 1996). An agent forms a goal-directed system with certain goals and actuation impulses, shown in the simplified Figure 3.2. Note that in the context of this diagram, “expression” refers to the intended expression of an agent. This differs from the actual observed behavior, which determines the observed expression by other agents. Force feedback, physical interactions, and automatic responses are part of the feedback loop noted between expression and behavior. Reflexes are one example of behavior which exists only in this feedback loop, without regard to goal state. This loop does not allow for meme learning but does greatly affect behavior.

Behavior does not perfectly reflect impulses or goals. The perceptual system can provide incomplete or incorrect information. Differences between the perceived world and the actual world provide sources for variation in reproducing memes. Muscle actuation introduces behavioral variation as well. Proprioception and internal control systems for muscles do not operate precisely, creating random variations in reproducing similar behaviors (Schmidt & Lee, 2005). Individual differences create systemic differences in meme behavior across agents.

The block diagram in Figure 3.3 synthesizes a loop for an agent interacting with the physical world. The stages from perceptual and control theory are noted. For the purposes of meme transmission, consider this diagram as representing one agent within a society that is connected by a common environment. The processes described in this diagram can explain reproduction as defined in Section 2.1, but with higher granularity. Established theory of social learning and information flow will be overlaid onto this diagram.

Figure 3.3: Agent Information Flow



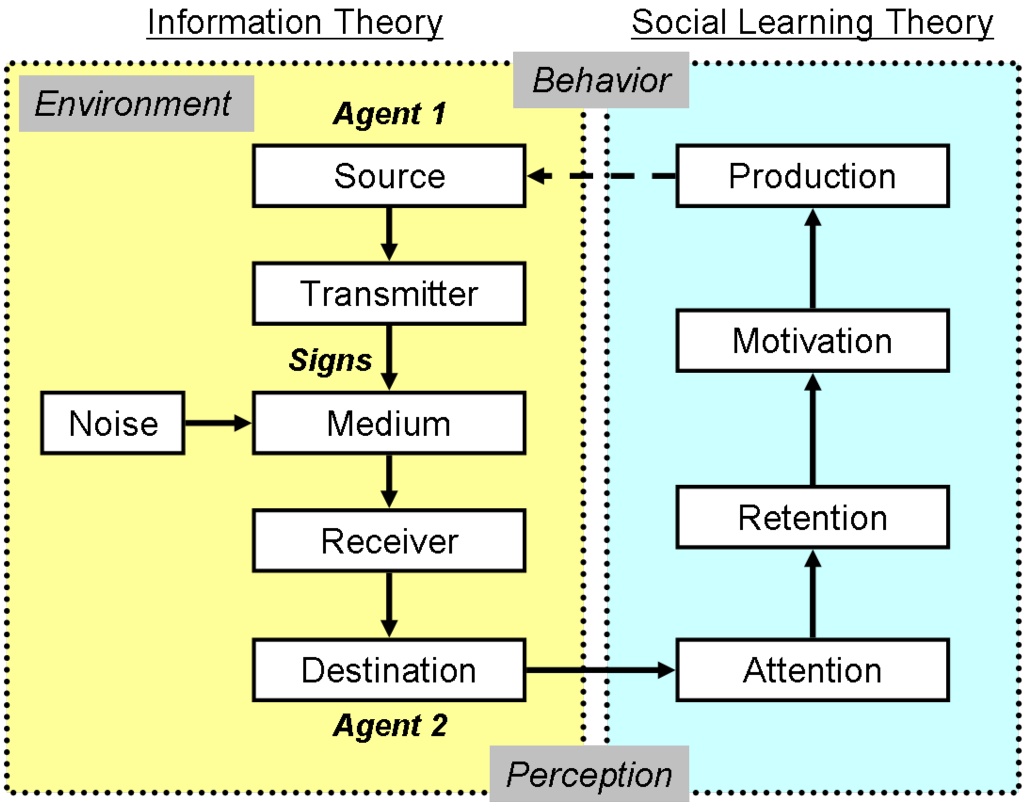
3.2 Synthesis of Models

Synthesizing social cognitive theory and Shannon (1948) information theory creates a systems model for meme reproduction. Figure 3.4 shows the processes involved in transmitting a meme from one agent to another. The dashed arrow in this diagram represents when processing by the original agent is complete and a transmission to a new agent can start. Information theory provides useful insight into the transformations of a meme's physical information while in transit: the changes to a meme from non-cognitive effects. Information transmission occurs between a source and a destination, also known as the target. The source chooses the content of the information, which is passed to the transmitter. The transmitter sends this message by altering some part of the medium, producing a signal. A medium is the channel through which information travels, such as television or speech. Noise affects the signal as it passes through this medium. The receiver monitors this medium and is affected by a signal, allowing the message to pass to the destination where it is interpreted.

The processes transmitter, medium, noise, and receiver occur during the "Transmission of Information" step marked in Figure 3.3. Social cognitive theory provides a framework to understand how a meme would be reproduced cognitively. Figure 3.5 overlays labels for the processes of social learning onto the information flow diagram of Figure 3.3. These theories complementarily address the meme reproduction process.

Social cognitive theory proposes a model for how humans learn information

Figure 3.4: Synthesis of Information Theory and Observational Learning

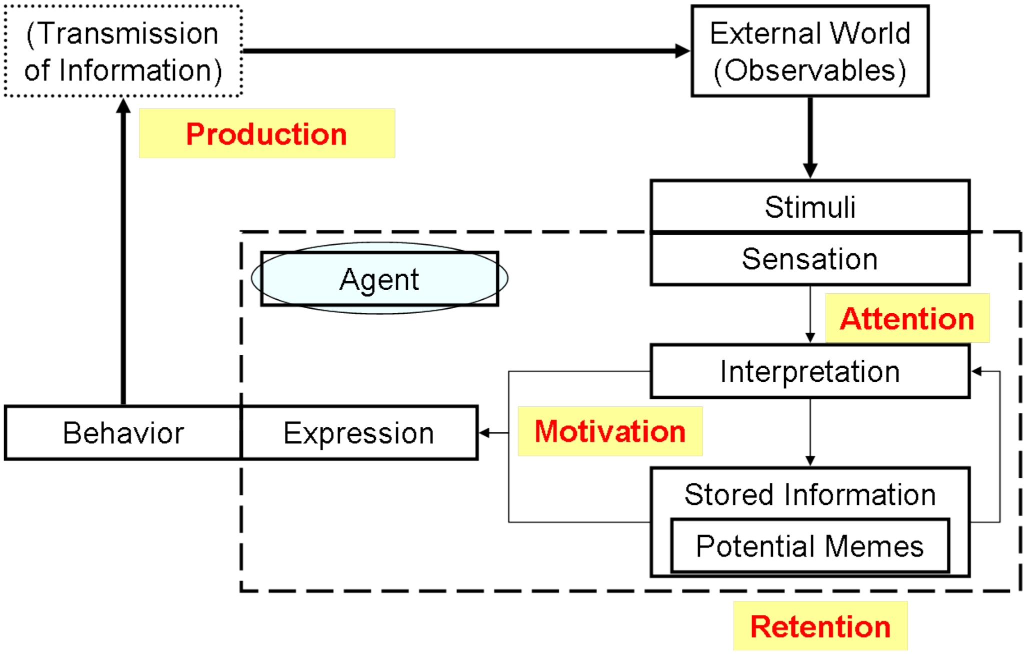


from other members of society. The precursor to social cognitive theory was social learning theory, which approached the problem from a behavioral view rather than a cognitive view. Social learning theory developed out of seminal findings of human imitation (Bandura, Ross, & Ross, 1963). The memetic process encompasses most kinds of imitation and emulation, making social learning theory a logical component of the system.

Social learning theory lacks explanatory power for the physical transmission of information. Shannon information theory explains fundamental processes of information transmission between a source and receiver through a medium. Information theory fills theoretical gaps that social learning theory does not address implicitly. Noise and bandwidth effects provide particularly important insight into meme mutation and selection pressures.

Social learning theory and information theory combine to create a system within which memes reproduce and evolve. Within this system a meme differs from other information only by its ability to fulfill the recursive reproduction condition (Eqn. 2.1). For certain societies, this space will be tightly constrained. For songbirds, only new songs are learned socially (West & King, 1996). For

Figure 3.5: Agent Social Learning Loop



humans, this space could reasonably include most observable behaviors.

While humans are biologically able to socially acquire a vast variety of information, only a subset of information disperses through society. Some memes must be more fit than others. As in biology, the fitness of an entity depends on its environment. Understanding meme fitness requires a deeper examination of the reproduction process in relation to a full society.

3.3 Information Theory View of Memes

The physical transmission of a meme interacts with noise and competes with alternative signals in the environment. Information theory must be applied to understand the transmission of memes through signs and behavior. Shannon's (1948) theory addresses how information must be sent, without addressing why or when an agent would send information. Information theory provides a framework to analyze the effects of receiving information, sending information, and noise.

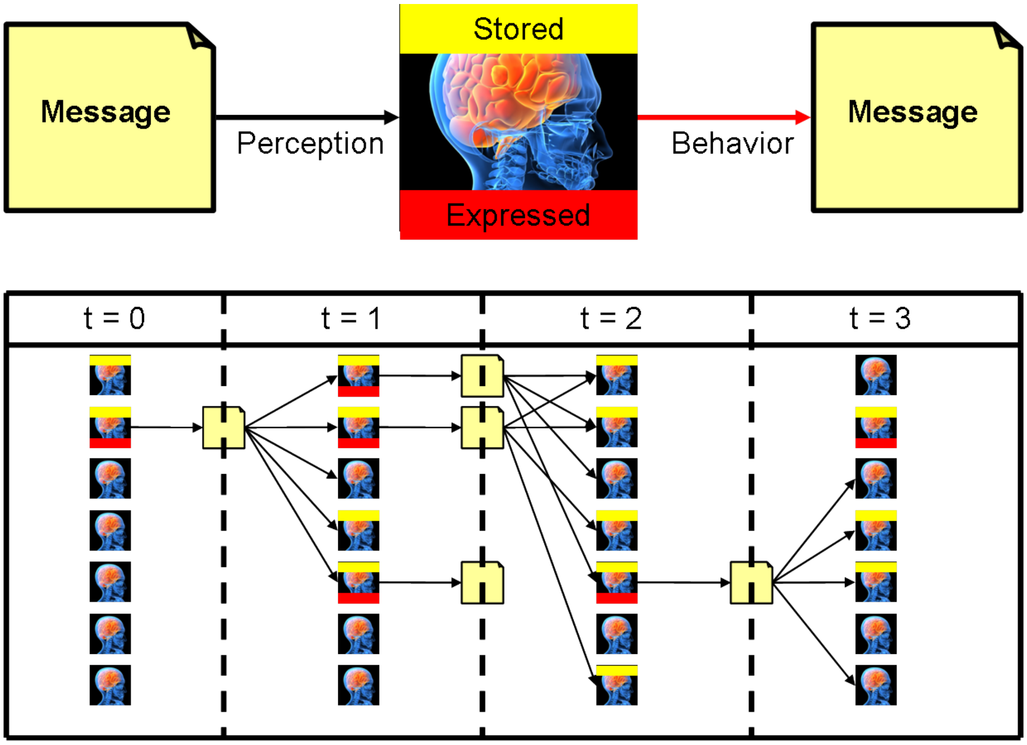
A meme transmission is a subset of generic messages within information theory. Meme transmissions do not act as dyadic transmissions. A source agent expresses a meme as behavior but may have no intent to communicate with a particular agent or any agent at all. Behavior broadcasts the meme into the environment, a physical medium for transmission. Within that medium, other mediums may exist, such as written language and physical demonstration.

Complicating matters, all agents actively participate in behavior and broadcast competing signals.

Agents in a static environment with unlimited perception, memory, and behavioral capabilities could spread information unencumbered by limitations on information transmissions. For realistic analysis, limits exist on agents' capabilities. Real environments dynamically change even in the absence of any behavior. Noise and bandwidth effects impede and break the transmission of memes.

3.3.1 Behavior as Transmission

Figure 3.6: Behavioral Transmission



Agents output behavior, which transmits information. Memes are a special type of information, which is transmitted and rebroadcast by new agents over time. Figure 3.6 shows how transmission might pass memes to new agents over time. Limited control and rate of activity bottleneck the information transmitted. An agent's control over the environment consists of their available outputs, such as muscle control and vocalization. An agent only controls a subset of behavioral inputs to the environment. The rate of output transmission also holds importance. For example, a faster writer can transmit more information per unit time. The product of the output vector and the transmission rate determines an information-theoretic bound on transmission rate of memetic information.

Behavior provides the opportunity for noise in transmission. Imprecise control of behavior introduces variation in implementation. For any behavior, physical implementation varies based on the actor and circumstances. Behavior expressing a meme may constitute a subset of an agent's total behavior, creating the potential for interference in transmission. Representing behavior as B_a, the effects of random variations (ϵ_a) and other activities ($\tilde{x}_a$) are explicitly represented in Eqn. 3.1.

$$\begin{aligned} Env^* &= B_a(Env, X_a) \\ &= B_a(Env, x_a + \tilde{x}_a) + \epsilon_a \end{aligned} \tag{3.1}$$

Behavioral bandwidth for transmitting information is limited. Behavioral interference indicates that behaviors compete for outputs, as might occur when multitasking. For example, speaking and eating a sandwich are semantically orthogonal but use the same muscle set. Mixing these activities degrades the performance of both.

3.3.2 The Dynamic Environment

Once transmitted, a message enters the environment through some medium or mediums. The environment interacts with messages dynamically through natural physical processes, introducing random and systemic noise. Research in linguistics (Dell, Chang, & Griffin, 1999), marketing (Costley, Das, & Brucks, 1997), and communication (Schramm, 1963) examine the implications of different message encodings and mediums on messages. All results pertinent to generic physical transmissions must apply to memes, which comprise a subset of messages.

The field of message passing cannot be done justice in a short section, so only the most fundamental processes will be examined. For a full treatment, Heath and Bryant (2000) provides an overview of transmission processes and protocols. Mediums have limited transmission rates and can superimpose dynamic noise onto a message. Message size, message fidelity, and signal to noise ratio significantly affect the variation and replication ability of memes (HaleEvans, 2006).

Message size influences the ability of a meme to propagate. Longer messages require greater transmission and reception times. The complete text of Hamlet reproduces more slowly than the phrase "To be or not to be." Long transmission times utilize more behavioral resources. This can reduce the opportunity and motivation to repeat a message. Production of signs such as books and mass media mitigate the behavioral constraints, as a single reproduction lasts longer.

For longer memes that cannot be expressed through a lasting medium such as print or photography, transmission can be extremely difficult. The difference between choreography and notational music highlights this issue. Sheet music preserves the sounds of composers such as Mozart and Bach, which can be expected to last for hundreds of years. Choreography lacks a universal format for recording movements, forcing complex dances to be taught through direct instruction or video presentation (HutchinsonGuest, 1989). This makes copying long sequences of choreographic movements time consuming, meaning that choreography is often lost once it is no longer performed (J. Anderson, 1982).

Different mediums vary by the characteristics of the noise introduced. Degradation (ϵ_d) and competition (ϵ_c) introduce noise into a signal through different mechanisms. Assume that the meme transmission carries the signal as a function TR of the meme information x. Equation 3.2 expresses the received message (R_a) by an agent a. All processes operate over time, expressed as t.

$$R_a(t) = TR_x(x,t) + \epsilon_d(t) + \epsilon_c(t) \tag{3.2}$$

Degradation reduces message quality due to entropy and other changes to the medium. Degradation differs for written and spoken forms of words- the waveform of a spoken word loses coherence more quickly than a printed word. The rate of degradation affects how long signs continue to present the physical information of a meme. Degradation may not represent a fully random process. Writing your name in the sand by the ocean will certainly degrade overnight, but not uniformly.

Competition introduces non-random noise that carries meaningful semantic information. Competition within a medium increases noise by superimposing an alternative signal. Competition within a medium can directly affect the physical value of a signal. Superimposing analog electrical signals creates a new signal, which may cancel out pieces of the original signal. Separate signals can be disentangled by relying on orthogonality of characteristics and error correction. FM radio employs this principle (Carlson, 1981). Simultaneous voices speaking follow this pattern.

Signals may combine such that the original signals are irretrievable. In this case, competition destroys components of one or both original signals. Changing the voicemail message on a phone deletes the old message, for instance. Entropy created by a large number of competing signals can also create this effect by introducing enough small errors as to be irrecoverable. A sufficiently large number of voices in a room can create this effect.

Noise and error correction introduce variation for memes. One criticism of memetics contends that this causes memes to vary too much to retain meaning (Atran, 2001). Considering that a word in language can be treated as a meme, this criticism lacks general appeal. Research on information theory has shown that variation can be reduced through messages with less information per length and coding methods such as parity. Discretized transmissions also have advantages over continuous transmissions; they reduce the set of allowable states in a signal. However, the introduction of variation does require careful definition of a meme for empirical study- a point addressed later in Section 4.1.

Degradation and competition are two sources of noise for a meme. The signal to noise ratio has important implications for a meme. Meme transmissions will encounter noise and may undergo error correction, depending on the medium and agents involved. Error correction and handling of incompletely received messages is a significant issue which will affect some memes, but it appears unclear what theories of human cognition imply for memes. Related work in linguistics (completing words and phrases) and visual perception (picture completion) has implications for this, but these are not sufficient. The most difficult question on this matter appears to be how information is handled when it is transmitted in multiple partially observed sessions. To this researcher's knowledge, a consensus mechanism of this mechanism does not exist. Accordingly, this conceptual model lacks a specific theory for handling incomplete meme transmission, forcing the problem to be addressed anew for any meme where this mechanism significantly influences dynamics.

Stimulation as Reception

Stimulation determines what environmental stimuli an agent physically detects through its bandwidth of detection channels. Sensory organs provide the set of stimuli for an agent and are necessary to receive a meme. While important biologically, this mechanism holds limited value for establishing the fitness of memes within a society. The sensitivity of sensory organs can provide insight into sources of noise and sensory error correction mechanisms. Significant empirical work has mapped out minimal detectable stimuli for human sensations, such as light detection and weight approximation (Levine & Shefner, 1991).

Being able to physically sense the signs of a meme provides a limited form of pruning out transmissions. While the ability to sense stimuli is necessary, sensation alone is not sufficient. Stimulation provides no insight into the success of a meme with regard to its semantic information. Failing to learn the lyrics to a song because the pitch fell outside one's hearing range provides meager insight into the workings of society.

Detection of stimuli conveying a meme is a necessary condition for meme transmission. This condition is somewhat lax, however. A variety of modes of transmission may exist and only one viable transmission mode must exist. The syntax-semantics divide allows for multiple representations of the same content (Halliday, 1978). Practically, this means that a possible meme can be ruled out only when the meme cannot be transmitted using any combination of available behavior and perception. Boundaries separating agents, such as spatial distance or language barriers, create this effect. Behavior, memes, and social factors will control some of these boundaries. For example, free individuals choose the context and persons they interact with. For the first order examination, these barriers can be taken as external to the model. This is because at any particular point in time, the physical barriers operate independently of their cause.

When agents interact, even in a limited manner, disproving the possibility of a physical transmission requires the proof of a negative: that no syntactic representation can spread the semantic information socially. Incommunicable learning such as the experience of color (qualia) and muscle memory exist, but these cases approach questions outside the useful scope of meme analysis. While it is important to establish conditions which debunk the existence of a meme, sensory constraints tell little about the vast possibilities for social learning in a human population. At the first order, physical barriers are a straightforward filter on the ability to detect a meme.

3.3.3 Evolutionary Mechanisms

When looking at memes, transmission of a meme supports inheritance while noise introduces variation. The balance between these mechanisms guides a meme's evolution. A perfectly reproduced meme cannot evolve, a competitive disadvantage. Conversely, a potential meme with a poor signal to noise ratio may totally lose the original signal. The received signal might or might not constitute a meme. Even if the received signal was a meme, no meme would have reproduced since the new meme did not inherit from the old meme. The game "Telephone" exemplifies this effect, where a message passed along a ring of players seldom bears any semantic resemblance to the original message.

High information messages suffer more from noise. In information theory, a random message holds the most information because no pieces may be inferred from other portions of the message. Treating the meme as a message, three implications emerge. Firstly, a lower information meme should undergo less variation per length of message. Secondly, a meme which fits expected transmissions will be corrected more effectively and create less variation. Thirdly, memes may not transfer properly due to their similarity with another meme. In this case, correcting for noise introduced with the meme will convert it to a known meme.

Signal transmissions depend on the signal to noise ratio, noise characteristics, and correction process. Research on noise sources, signal quality, and encoding techniques provides insight which should be applicable in estimating the variation of memes in transmission. Productive research on these topics at the theoretical and practical levels can be found in diachronic linguistics literature, which examines the evolution of words (Labov, 1994).

Defining noise requires defining a signal, allowing analysis of the signal to noise ratio. The nature of noise reduction and error correction processes in organisms is not well understood, an open question relevant to memetics. From an agent's standpoint, a meme holds no intrinsic status as a "meme of interest." On the contrary, if an agent needed to expect a meme to learn a meme then it would not be very effective. Memes differ from other information because of how agents react to processing them, not because they have the ability to access some special meme processing mechanisms. How an agent attends to and acquires memes cannot differ from its normal perceptual process. The nature of perception and attention places bounds on the fitness of memes. Motivation to retransmit a meme also must play a role. Understanding these processes requires understanding the organism receiving a meme, which will be framed using social learning theories.

3.4 Memes As Social Learning

Meme theory must be considered as a recursive expansion of social learning, concerned with individual interactions for their influence on the cultural plane. Behavioral psychologists have identified a number of distinct mechanisms by which animals acquire behavior from other animals (Zentall, 2007). While not all memes involve imitation, every case of true imitation may be considered a meme. Appendix B notes the different mechanisms identified and if such learning constitutes a meme. Blackmore (1999) addresses imitation mechanisms in *The Meme Machine*, providing a complementary account for these distinctions.

Within the context of a single meme replication, social learning theory outlines a mechanism through which agents process and reproduce memes. The Bandura (1986) observational learning process proposes that humans pass knowledge through processes of attention, retention, motivation, and production. Attention is the process that filters signals out of the environmental medium. Retention is the process that stores a meme, allowing that information to influence behavior. Motivation is the process by which an agent chooses which behaviors to express, selecting between memes. Production is the process that produces behavior out of agents' goals. These processes dovetail with the effects of information theory, providing a system for syntactic and semantic transmission.

Observational learning defines cognitive processes connecting perception with behavior. Agents receive information through perception, with memes being one type of information. The key processes of interest in perception are stimulation, attention, and interpretation (James, 1890). Stimulation determines an agent's inputs, attention determines filters on inputs, and interpretation converges on a meaning for the configuration of inputs. These processes determine what *can be* detected (stimulation), what *is* detected (attention), and what the stimuli *means* (interpretation). Using perception and cognitive processes, an agent regulates their behavior.

Stimulation and behavior's role in memes has often been addressed from a deterministic stimulus-response (S-R) behavioralist perspective. This approach ignores the cognitive machinery that determines how and why a meme might be reproduced. The "thought virus" view of memes adoptions fosters this attitude, removing much of the importance of the agent (Dennett, 1995). It is a useful analogy but not a useful explanation. Social cognitive theory acts as a cognitive bridge between stimulation and behavior, providing a framework for interpretation of perception and its influence on behavioral patterns. Each process of observational learning has a key role in the reproduction of a meme.

3.4.1 Attention

Attention interacts with both the semantic content and the physical content of a message. Attention provides a process that solves the problem of which stimuli over time constitute a signal. To receive a meme transmission, an agent must attend to it. In this context, passive and active attention will not be differentiated; either may influence social learning.

Attention is a limited resource. Within a complex environment, an agent cannot process all signals completely. The cocktail party effect demonstrates that in processing speech, attending to one speaker significantly reduces attention to a second simultaneous speaker (Cherry, 1953). Attending to one signal can mean attenuating a competing signal. For this reason, attention will be treated as a signal filter. By treating some signal characteristics as noise, others may be analyzed coherently.

Two types of attention exist: early selection and late selection. Early selection attention filters based upon raw physical qualities (Treisman & Gelade, 1980). The bandwidth filter on a radio operates along this principle. Late selection attention filters based on semantic content contained in a signal (Deutsch & Deutsch, 1963; Treisman & Gelade, 1980). Attending to one's name in a noisy room exemplifies this effect. Early and late selection are not exclusive, with psychology and machine vision utilizing a combination of filters approach (Johnston & Dark, 1986; Backer, Mertsching, & Bollmann, 2001).

The Attentional Bottleneck

Due to the complexities of a real society, attention could be the primary mediating factor of meme reproduction. The likelihood of a bottleneck at this stage can be explained using information theory. This analysis ignores the effects of signal interference, but this effect should not change the intuition gained.

Queuing theory can be applied to generate an estimate of the total social information transmitted. Assume a society S of N agents, with each agent a participating in a similar behavior B_a. Signs of behavior are generated with some arrival rate λ_{B_a}. These signs exist for some variable length t_{B_a}, with mean duration μ_{B_a}. An $M/M/\infty$ queue approximation yields the average number of an agent's signs in the system, shown in Eqn. 3.3.

$$E[\sharp Signs_a] = \lambda_{B_a} \times \mu_{B_a} \tag{3.3}$$

N agents acting simultaneously will create N sets of distinct signs at each given point in time. Assuming an agent physically can detect all signs of behavior, the steady state number of signs in the system ($E[\sharp Signs]$) may be stated as Eqn. 3.4. This expression sidesteps the effects of interference, which might be represented by shorter average durations. This amount of information does not even account for attention to internal feedback for behavior or non-social information within the environment.

$$E[\sharp Signs] = E[\sum_{a}^{N} \sharp Signs_a] = N \times \lambda_{B_a} \times \mu_{B_a} \tag{3.4}$$

Behavior in progress provides one form of sign, so the term $\lambda_{B_a} \times \mu_{B_a}$ will always be greater than one. The number of available signs will be no less than the number of other agents and could be much greater. While the mean duration of a sign such as a blink or a clap may be relatively quick, written signs and constructs survive for generations. While not all signs may be present at any one time, they all exist in the environment and could potentially be perceived. This accumulation leads to a vast number of signs.

In a society of five, each agent would need to perceive information at least four times as quickly as it transmits in order to attend to all the other agents. Sign language disappears nearly instantaneously after it is presented, staying near this lower bound. For written language, μ_{B_a} can be two or more orders of magnitude higher than λ_{B_a}. An agent would have to be able to attend all the books written by other agents while simultaneously watching the next books being written. Unless an agent's sensory capabilities take in very little information, processing every signal will be infeasible and wasteful.

The given example of reading multiple books fully as they are written could leave you scratching your head. Why would an agent re-read an entire book instantaneously for every second of its existence, even if it could? The novel would no longer be novel, so to speak. For an agent that learns and remembers, a single reading could be sufficient. Novelty is factor that influences attention. However, novelty is only one of many characteristics affecting information for processing.

Attentional processes should facilitate processing relevant stimuli (Wickens, 1991). Relevance of information depends on its impact on an agent's goals and needs. Normatively, attended signals should be ones that improve an agent's ability to complete their goals. Attenuation theories propose that unattended stimuli are tuned out, freeing cognitive resources to process the remaining inputs (Eysenck, 1982). Identifying relevant stimuli and events poses a non-trivial problem.

Attentional Salience

The true relevance of a signal cannot be known, so theories of attention refer to the salience of a signal (Fazio, RoskosEwoldsen, & Powell, 1994). Salience may be considered a heuristic for relevance of information for an adaptive agent. Within this discussion, salience will not refer to the Bandura (1986) view of salience as the properties of an action. Instead, this discussion uses the broader concept of perceptional salience as used by William James (1890). Psychology identifies signal quality (Posner, Snyder, & Davidson, 1980), novelty (James, 1890), motivation (Mogg, Bradley, Hyare, & Lee, 1998), selection (Hastorf & Cantril, 1954), duration (Shibuya & Bundesen, 1988), and frequency (Lee, Itti, Koch, & Braun, 1999) as causal factors in mediating attention. Other factors have been proposed which have merit but will not be discussed due to space constraints. Pashler (1998) gives a more complete overview of attention literature.

Signal quality commonly refers to metrics such as signal to noise ratio. From the source's standpoint, signal quality indicates the preservation of a transmitted signal. For the receiver, the attentional process must determine which signals constitute noise. Defining signal quality in these terms would be circular. The perceived quality of a signal will be considered the ability to parse semantic information from a signal (Oakley, 2007). A lower quality signal indicates a less comprehensible signal, impeding assignment of meaning. Less comprehensible signals appear to receive less attention in humans and other animals (D. R. Anderson, 1981).

Novelty and surprise lie at the heart of attentional theories based on learning. In humans and animals, novel information receives more attention (James, 1890). For an agent capable of pattern memory, unchanging patterns no longer provide new information. For a learning agent, novel information improves awareness of the environment.

Signal quality and novelty each support two separate factors, a syntactic component and a semantic component. A light that begins flashing may be physically novel but it is devoid of semantic information. Conversely, the flatline signal for a patient in a hospital provides no syntactic novelty but its continued presentation signifies the onset of death. Table 3.1 indicates the different types of factors. The syntactic and semantic components will be assumed to act similarly and independently, though evidence on this matter is not conclusive.

Table 3.1: Semantic and Syntactic Attention

	Syntax	Semantics
Novelty	Contrast	Unfamiliarity
Signal Quality	Separability	Comprehensibility

It should be noted that novelty and quality can work against each other. A signal consisting of incomprehensible noise holds the most novelty, containing fully unpredictable information (Shannon, 1948). Noise holds the most syntactic information and the least semantic information. Bandura (1986, p.59) posits, "Retention improves by transforming the meaningless into what is already known," indicating that familiar information will be more easily received. Filtering out signals from these extremes forms an adaptive balance, attending to information that has meaning without being predictable.

Motivation indicates a goal or need state for an agent. Higher motivation increases the salience of associated stimuli (Fazio et al., 1994). For example, hungry individuals detect food cues more readily than controls (Mogg et al., 1998). From an evolutionary standpoint, this increased salience seems readily explainable. Despite the obvious impact of motivation, measurement and understanding of the processes that engender motivation are not fully understood. Motivation will be the focus of Section 3.4.3.

Selection refers to a signal currently undergoing processing. Signals attended at the current time will be more likely to maintain attention, giving them a higher salience (Eysenck, 1982). A selected signal has an advantage over unselected signals, as these receive differentially less processing. Inattentional blindness demonstrates this effect, where selective attention blocks out otherwise salient stimuli (Simons & Chabris, 1999).

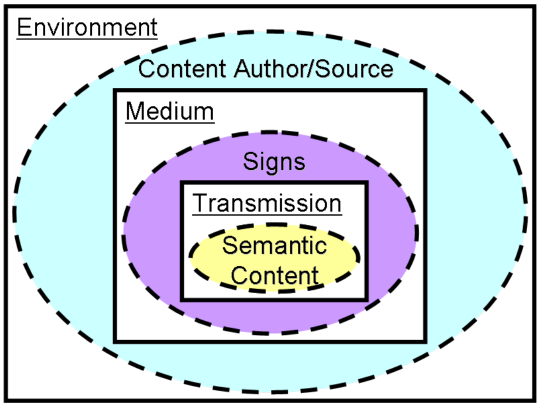
Duration and frequency of presentation have a contextual value for mediating attention. Increased duration of presentation gives more time for attention to fixate on a stimuli (Shibuya & Bundesen, 1988). Increased frequency of presentation provides more opportunities for a stimuli to receive attention, reflecting a greater total level of contrast provided over time (Ray, Sawyer, & Strong, 1971). Habituation works against these mechanisms, reducing the novelty of a stimulus after many presentations (Rescorla & Wagner, 1972).

These factors have a combined effect on salience which is not well known. This problem is endemic to experimental study, which is typically designed to control for the maximum possible factors in order to maximize effect size. As a result, the covariance between factors is not well understood. Within the larger behavioral system, these gaps in knowledge expose avenues for further research which necessitate modeling assumptions and calibration when simulating memes, as discussed in Chapter 5.

Attention to Sources

The syntactic and semantic content of a meme's signs must be filtered by attention. However, memes are socially transmitted information; they imply a source. Due to selective attention toward different sources, memes transmitting from attended sources have a strong advantage. Bandura recognized this influence, focusing on the source as the "model" for behavior (Bandura, 1986). In society, sources do not just consist of agents. Signs such as written accounts and recordings also transmit memes.

Figure 3.7: Scopes of Attention



Communication theory approaches these issues (Schramm, 1963). Three syntactic scopes exist which require attentional filtering: environment, medium, and transmission. These scopes have a containership relationship as presented in Figure 3.7. Semantic information is associated with each level: the content source in the environment, the signs in the medium, and the semantic content of the transmission. The existence of different scopes for attention indicates a hierarchy of attention needed to receive a meme transmission (Oakley, 2007).

For example, imagine a crowded poster session. Within the environment of the room, posters (mediums) must be identified before their words (transmission) can be read, and words must be recognized before deciding which words to read. Semantic cues guide this process. Posters presented by famous renowned authors (content sources) will draw greater attention, as will posters featuring more familiar language (signs) and interesting content (semantics).

From the cognitive standpoint, only assumed and perceived semantic cues matter. Semantic cues may be missing or ambiguous. In particular, the source of content often involves significant ambiguity, sometimes intentionally. Semantics can also interact: a bad idea from a brilliant person makes the person seem less brilliant. The grounding model of cultural transmission formalizes this interaction, proposing that message content establishes identity and identity changes the meaning of content (Kashima, Klein, & Clark, 2007).

Perceptual processes of segmenting the environment, defining perceptual events, and receiving semantic information each involve attentional processes (J. J. Gibson, 1986). The tiered nature of attention has major implications for the spread of memes, especially in an agent that can freely select its interactions. In order to receive a meme, an agent generally must recognize and attend to its source. Bandura (1986, p. 54) states, "People are more likely to select models who are proficient at producing good outcomes..." and "Models who are interesting and otherwise rewarding tend to be sought out...."

Social psychology research supports these hypotheses for human interaction (Kelley, 1955). Similar results occur when different source characteristics are attached to otherwise identical messages (Chaiken, Wood, & Eagly, 1996). Trademarking and branding within marketing attempt to increase attention to a source of products and information. Social psychology and marketing effects apply to both attention and motivation processes. Specific social psychology influences on meme transmission are postponed into the discussion on motivation in Section 3.4.3.

3.4.2 Retention

Once a meme has been attended to, there must be an effect on the agent's memory in order to retain the semantic information in some form. Retention of social learning necessitates some storage of the semantics related to a meme, which can later be reflected in their behavioral patterns. Bandura (1986) concentrates significantly on the symbolic representation of social knowledge. The social cognitive theory focuses on this matter in a manner more specific than can be applied to memes in general, concentrating on the representation of action sequences.

From the meme standpoint, retention matters with regards to how different semantic information can be retained and recalled. The effects of limitations on retention, incomplete retention, and addition of contextual information on memes could provide insight into some mechanisms for variation and selection. Unfortunately, confounds exist for understanding of retention. Memory representations cannot be directly measured or mapped (Uttal, 2001). Brain localization techniques such as fMRI mapping approach this problem, but currently cannot measure knowledge or infer the semantic web of an individual.

Retention cannot be measured except through changes in behavior. Brain imaging experiments require baselines and so can only study retention when combined with a learning task. This means that retention, motivation, and production of behavior cannot always be easily disentangled. For example, proving that organisms forget information has been exceedingly difficult because there is no way to separate forgetting from an inability to retrieve stored data (Bjork, 2003).

Research on education and memory has implications for retention of memes. In particular, the effect of scaffolding on social learning has had a measurable effect on the ability to reproduce behavior (HmeloSilver, Duncan, & Chinn, 2007). Scaffolding posits that skills and information build off known information (Vygotsky, 1980). This point was echoed by Bandura, as cited in Section 3.4.1. For certain knowledge, this effect has intuitive value. Ten years of attending advanced calculus lectures will not benefit a student as much as a sequenced curriculum starting with basic math skills. For other types of knowledge, such a relationship may not hold. For example, languages can be effectively taught through immersion techniques (Cummins, 1998).

For limited cases involving the ability to learn visible behaviors, social cognitive theory provides a framework for understanding retention and refining of stored semantics. For memes in general, the effects of retention on fitness and variation depend on open questions currently being explored in semiotics and psychology. It is possible that different types of information could be even stored by qualitatively different mechanisms, meaning that the content of a meme influences retention effects. With this stated, some implications from learning research will be examined for their implications for memes.

Common findings on retention incorporated into learning theories are the number of presentations, usage of information, and parsing into semantics. The number of attended presentations increases the likelihood of recall and accuracy of recognition (Hintzman & Block, 1971). Repetition of information tends to improve retention as well, forming the basis of rote learning. Understanding an observed event, rather than rote memorization of physical stimuli, also helps retention (Jacoby & Dallas, 1981). Mnemonic devices provide a method for adding semantic layers to arbitrary information to harness this effect. While seemingly obvious, these three factors probably account for much of the variance in retention.

Surprise-based learning theories such as the Rescorla-Wagner model posit that learning fits an error correction model (Rescorla & Wagner, 1972). The least expected information provides the highest magnitude of learning in a standard error correction model. Emotional tagging also appears to play a role in the encoding and recall of memories (Canli, Zhao, Brewer, Gabrieli, & Cahill, 2000). Reinforcement learning approaches utilize similar error-correction principles, but with regard to rewards for behavior (Kaelbling, Littman, & Moore, 1996). Models of this sort would predict that memes which violate existing knowledge would be remembered better. Evidence from cognitive biases and persuasion research conflicts on this matter, with the confirmation bias showing a preference for expected information (Nickerson, 1998) but depth of processing metrics showing additional processing of unexpected information in certain circumstances (Jain & Maheswaran, 2000).

The Rescorla-Wagner model also exemplifies associative learning. Associative learning posits that learning consists of associations created in the mind (Mackintosh, 1983). Stimulus-response (S-R) models commonly assume a single layer of learning, while connectionist models consider a web of associations such as a semantic net. The implications of associative learning structures on meme reproduction are unclear. Associative learning implies connection strengths between mental constructs, but the nature of the constituent constructs is unclear. For associative learning to provide insight into memes requires assumptions about the set of mental constructs, which cannot be readily verified.

Conversely, models such as trained neural nets avoid defining constructs but depend on their precise configuration of weights (Hebb, 1949). This level of analysis appears too fine grained for examining memes given the current state of the art. By depending on combinatorial sets of weights, connectionist approaches provide limited added insight into the level of retention beyond that from reinforcement learning. Where a concept association model might require assuming concepts and connections, a neural net requires assuming the number of neurons and layers plus a myriad of weights with unclear meaning. This does not mean to imply that neural network models of the mind are incorrect; they aim toward an analogue of the human brain. However, they are limited in their insight for memes. Neural networks may be useful for determining the maximum complexity of a meme to be retained, treating them as patterns. They also offer mechanisms for the degradation of memory patterns through washing out of weight configurations (Sikstrom, 1999). It is possible that simpler mechanisms could be employed to attain the same insights, however.

Schema learning theories posit two processes of retention that are of interest: assimilation and accommodation (Axelrod, 1973; Piaget, 1955). Schemas may be thought of as abstract patterns, such as the general qualities defining a tree. Assimilation involves fitting a new stimulus into an existing pattern, potentially updating the pattern. Another type of retention in a schema pattern, accommodation, involves defining a new schema. Differentiation is one type of accommodation, which defines a new schema by its differences from known schemas. Differentiation might occur when learning which size objects are too large to grab with one hand, providing a rule to segment the world into what can and cannot be grasped. Eleanor Gibson has outlined plausible processes for perceptual learning based upon differentiation, making it a potential mechanism for other forms of learning (E. J. Gibson & Pick, 2000). Accommodation involves updating an existing schema based on new information.

Learning through these schema mechanisms would imply a non-linear retention curve for memes. Memes too similar to an existing meme would be absorbed through the assimilation process, potentially losing their unique information. Conversely, a meme completely unrelated to any known information might have no base from which to differentiate and be lost. These competing factors have some similarity to the balance between novelty and content noted when discussing attention in Section 3.4.1.

Schemas capture similar elements as discussed by Bandura (1986) as symbols. They also dovetail with a concept-oriented associational approach. While theoretically appealing, schemas bring significant assumptions which may not be justified. Research has not proved the existence of unique schemas within the brain. Multiple schemas could exist independently for the same semantics, disconnected and assimilating separately. Accommodation might not occur at all, with each new piece of information retained uniquely (Sadoski, 1991).

The effects of retention on meme reproduction stymie clear solutions. Fortunately, the implications of retention may not be as problematic within a real society. Firstly, the vast memory capacity of a human potentially allows for a proportionally vast number of memes. Over the long run, common factors such as the number of attended meme expressions probably account for the most variance. Finally, memes of interest have the potential to be discussed and remembered with some fidelity. If they could not, it would be exceedingly difficult to study them academically- one scholar could not readily pass the concept to another scholar.

3.4.3 Motivation

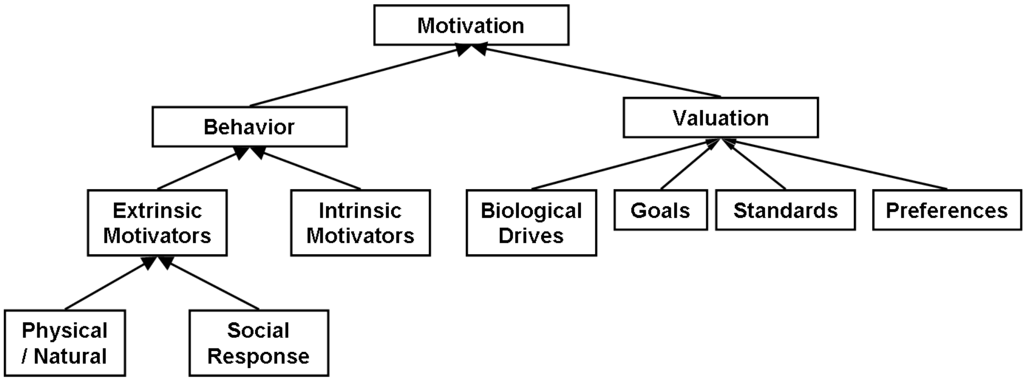
Motivation determines which memes an agent will express. Along with attention, motivation must be considered a key factor in determining the spread of memes. Returning to the definition of memes from Section 2.3, motivation determines the environment, if any, where an agent will express a meme. When and where an agent expresses a meme can have vital consequences for its transmission.

Consider this meme related to the tragedy of the commons. Assume one person discovers a way to anonymously steal food from the town granary, but the people in town might still learn the process of the action from the evidence left behind. The innovator has strong motivation against using this method in contexts where other agents could learn the process, as widespread theft would deplete the granary. As more people learn the practice, the motivation to hide its details would decrease and the meme could spread freely. Motivation effects might be particularly important for intellectual property discussions. For example, the Coca Cola recipe fits this pattern: its value as a brand depends on the secrecy of the recipe and the motivation for secrecy depends on the value of the brand.

Stronger perceived incentives toward certain behaviors will influence decision processes and even unconscious behavior. Psychology and economics literature both address motivation, unfortunately with only limited dialog between them (Bruni & Sugden, 2007). Research from each field will be applied to examine how motivation mediates meme expression and how memes change motivation. Motivational influences on memes refer to why an agent might express a meme. Changes to motivation encompass areas such as learning, attitude change, and persuasion.

Agents have differential motivation to express memes as a function of the surrounding environment and their mental state, such as goals and affect. Motivation may be split into the "what" and the "why": motivators and valuation. Motivators consist of states or changes which may be internal, such as pain, or external, such as receiving money (Maslow, 1943). Valuation indicates a cognitive process which determines the level of activation for behaviors as a function of their associated motivators (Sugrue, Corrado, & Newsome, 2005). Figure 3.8 breaks down motivation into components. These components are integrated upward, forming the agent's total motivation to express a meme. The left branch notes properties of a behavior and its results that may be processed as motivators, such as physical effects (e.g. outcomes). The right branch indicates personal factors about an agent that would affect their valuation of these motivators (e.g. desirability of outcomes).

Figure 3.8: Components of Motivation



Economists analyze incentive structures, providing a body of literature relevant to motivation. Decision theory examines how motivation and available options map into behavior, with utility theory most commonly explored. Utility functions may be considered a measure of motivation under this definition, where the inputs are motivators and the function calculates a valuation (Bell, Raiffa, & Tversky, 1988). Valuation need not provide a calculable intermediate value, however. Neural nets and decision field processes can perform robust error correction toward goal states without an explicit calculation of motivation value (Van Gelder, 1999). Intermediate values such as utility can be convenient for analysis and conceptualization, but this paper will avoid treating them as describing organic thought.

Bounded rationality conditions indicate that no person could explicitly weigh and compare all possible behaviors available to them (Simon, 1982). Even if such a calculation was possible, the level of waste would be evolutionarily disadvantageous. Agents simply do not weigh each behavior by enumerating its motivators, applying a value function, and selecting the highest. Even when humans explicitly attempt multi-attribute comparison of alternatives, they tend to leave out attributes and show poor internal consistency (Saaty, 1996).

The lack of explicit value calculations and internal consistency raises questions about the existence of motivators as anything but research constructs. This is not a fatal flaw to rationality, as the lack of rational behavior for a single decision may be the result of heuristics that are boundedly rational over many decisions. Such heuristics will depend on activation of behavior based on internal state and external stimuli. If an agent activates behavior based upon a combination of internal state and external stimuli, why must a set of consistent motivators exist across behaviors? This problem is related to the questions surrounding retention raised in Section 3.4.2, relating to the cognitive structure of motivation.

Schema theories and generalization principles offer support for common motivators across a variety of behaviors. In this context, reuse of motivators provides the ability to assimilate information about a motivator in one context and generalize this knowledge. A lack of connected motivators would cause overfitting of motivation, making behavior and motivators hard to generalize. For example, assume a contractor works two projects that pay in dollars and two projects that pay in euros. If the value of the dollar drops significantly and the contractor notes this for one job, they should generalize this knowledge to both jobs paying in dollars. Since changing the motivator for one job affects another job, a shared motivator appears likely. Not all motivators necessarily decouple from behavior, necessitating distinctions between motivators.

Motivators

Separating motivators from behavior allows classification of motivators. Psychology has analyzed motivations from a variety of standpoints including biological drives (Hull, 1943), need hierarchies (Maslow, 1943), behavioral tendencies (Premack, 1963), goal achievement (McClelland, 1976), and anticipatory self-regulation (Bandura, 1986). No single set of these factors fully explains motivation and ongoing research attempts to define new sets of motivators.

Depending on the semantics of a meme it can create a motivator, a change of valuation of motivators, or a new behavior that may imply motivators. Two approaches can be taken to specifying motivators, each with the goal of explaining variance in motivation to express a meme. For higher fidelity simulation, motivators may be optimized to explain the most variance. This approach might mean generating ad hoc motivators by fitting a training data set. For research, utilizing a set of motivators from literature has advantages. While an optimized set of motivators might better match a particular situation, comparing meme spread based on arbitrary motivators lacks theoretical appeal. Employing theory based motivators also generates feedback as to their explanatory power and shortcomings.

The social cognitive theory provides a composite view of motivation which will be used as a starting point. Bandura (1986) categorizes motivation by its vicariousness, externality, attribution of control, cognitivity, sociality, and probability. These motivator classifications mix the payoff properties of a motivator (externality, sociality, cognitivity), the potential results of behavior (attribution of control, probability), and the method by which the motivator was learned (vicariousness). A final factor not deeply addressed is contingent planning, surprising considering the social cognitive theory focuses on anticipated consequences. Table 3.2 notes parameters likely to influence meme expression.

Table 3.2: Factors of Motivators

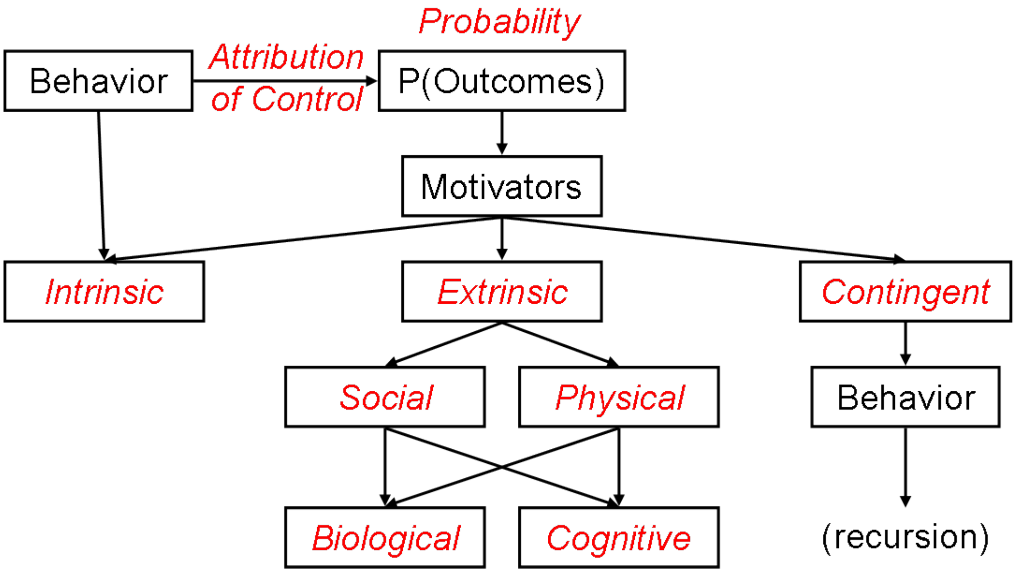
Factor	Converse	Description
Externality	Intrinsic/Extrinsic	Environmental outcome vs Enjoy act
Sociality	Physical/Social	Mechanical laws vs Social response
Cognitivity	Cognitive/Biological	Goal achievement vs Biological drive
Control	Internal/External	Able to control vs Powerless
Likelihood	High/Low	Probability of this outcome
Contingencies	Plan/No Plan	Enables more behavior and motivators

Figure 3.9 represents these factors in a decision theory format (decision, possible results, possible payoffs). An agent's internal state and environment could affect any of these factors and should be considered implicit to this diagram. This conceptual representation assists discussion of each factor's influence on meme expression, but should not be taken as an accurate description of real human thought processes.

Payoffs of Motivators

The payoffs of motivators provide an interesting path for analysis from the standpoint of memes. As noted, a major question for memes is why certain agents reproduce memes while others do not. This question of cultural and individual

Figure 3.9: Decision Theory Mapping of Motivators



differences can be approached by examining why agents value different behaviors and outcomes. Externality of motivation has implications for the stability of meme expression across contexts. Sociality of motivation determines if the physical or social environment will most influence meme expression. Cognitivity indicates the strength of experience over nature on meme expression.

While this is an important topic that gives rise to many interesting concepts and potential experiments for meme transmission, Section 3.4.3 on persuasion addresses the basic discussion of motivators necessary for understanding the mechanisms of meme transmission. For additional information, see Appendix D for a detailed discussion on the theoretical implications of motivator types.

Expected Outcomes of Behavior

An agent's self-evaluated ability to translate their intent into outcomes will influence the expression of memes (Bandura, 1986). Two sources add noise in mapping agent intent into outcomes. Firstly, behavior may have only a limited or unknown effect on the distribution of possible outcomes. This may be due to bounds on the precision of action, lack of information about the system, or constraints imposed by the environment. The attribution of control for an agent represents their perception of control over this distribution. Secondly, the distribution of outcomes may be probabilistic or incompletely known. This may be due to lack of information or fundamentally random elements of the system. The perceived probability represents the likelihood of a particular outcome from a selected distribution. Each factor has a role in mediating meme expression.

There are a variety of ways of looking at the effects of probability and expected outcomes. There is significant evidence that human handling of probability is non-normative and often fairly limited. While perceptions of control, risk, and uncertainty will have effects, these are not a primary focus of this paper. Accordingly, detailed exploration of psychological mechanisms related to these effects has been included in Appendix C.

As an alternative perspective, the somatic marker theory posits that decision making occurs via two qualitatively different processes- one being a filtering process and the other comparing outcomes based upon a form of subjective utility (Damasio, 1994). In this view, a significant portion of motivation and selection of actions would be driven by emotional tagging that would be highly specific to the individual. This would have some significant implications for the motivation to express a meme, being a sort of dual-process decision model. In particular, this sort of theory implicates that emotional tagging forms a significant connection between retrieval and motivation. From this perspective, the expected outcomes may be far less important than the emotional tagging from when a meme was learned and observed.

As such, there is significant active debate as to the relative importance of prior emotional experiences versus the rational analysis of the expected outcomes. In a respect, these are two different paths to a common destination. So long as past emotional tagging is a good indicator of future outcomes, both approaches should be expected to produce similar decisions. The two approaches diverge if the outcomes for a decision will be significantly different than prior experiences.

The workings of this mechanism are important to understanding the motivation process behind choosing to express memes. Somatic markers and other experiential context cues strengthen the influence of peripheral cues: context which is not directly related to the semantic content or outcomes of expressing a meme. In theory, these could result in memes being reproduced entirely based upon peripheral characteristics. People might learn memes because of attention to a source and also express those memes entirely due to the positive markers introduced by the source. As such, social influence could affect attention, motivation, or both.

Dual process persuasion theories, discussed in Section 3.4.3, have differing opinions as to the nature and relative influence of central and peripheral information. From this scholar's standpoint, theories of emotional decision-making such as Damasio (1994) have strong parallels with theories of persuasion such as ELM (Petty & Cacioppo, 1986). As such, it seems likely that second-order effects exist which augment the attractiveness of memes independently of their semantic content or expression outcomes. Further research on motivation and decision-making could greatly clarify the role of these different influences on meme expression. While such second order effects are not considered within the computational implementation presented in Chapter 5, they are an important open question for considering the spread of memes.

Changes in Motivation

Feedback exists between memes and motivation. While motivation affects meme expression, learning a meme impacts motivation. Memes carry semantic information, forming a special type of persuasive message. Persuasive messages can induce attitude change, emotional states, and changes in identity (Wood, 2000). Memes that shift attitudes and values can act in a facilitatory or inhibitory manner for other memes. Blackmore (1999) uses this effect to support the memeplexes, an interdependent collection of memes.

If a meme generates new goals or attitudes, it has changed an agent's motivation. Several possible avenues exist for change of motivation. These avenues are not unique to memes but could technically occur due to any new information. Memes could change motivation by one or more of the following processes:

1. Creating/destroying motivators
2. Increasing/decreasing the importance of motivators
3. Associating/disassociating motivators with outcomes

Motivation researchers disagree about the degree to which motivators are dynamic. Herzberg, Mausner, and Snyderman (1993) and Reiss (2004) assume static sets of motivators, used to evaluate situations. In this view, memes would be incapable of changing the set of motivators. Other researchers believe that people learn new motivators by associating a non-motivator with a motivator, making it a secondary reinforcer (Delgado, Labouliere, & Phelps, 2006). A final view posits that motivators spawn arbitrarily due to cognitive patterns such as goals, without requiring association to existing motivators (Sloman, 1998). Under associative and cognitive perspectives, a meme can create a new motivator. Such a motivator would influence all related behavior.

Changing the relative importance of motivators slightly could lead to equally significant changes in behavior. Meme acquisition and expression have the potential to change the relative importance of motivators. Acquiring a meme could change the perceived worth of an existing motivator, such as a dieting meme decreasing the value of fine dining. Expressing a meme could also change the perceived worth of motivators, such as through familiarity effects (Ortony, Clore, & Collins, 1988). With respect to memes, motivators of zero weight can be considered to effectively not exist. In this way, changing valuation can emulate creation and destruction of motivators.

A meme could instead provide information which alters the association between a motivator with either a behavior or an outcome. In this case, the underlying value of motivators do not change but they relate to behavior differently. A meme stating that cell phones cause cancer works in this manner. While it changes neither the value of communication or cancer, associating cancer with cell phones reduces the motivation to use one. Compared to directly changing motivators, this mechanism works much more specifically.

Motivator changes may be considered as fundamental valuation changes, while association changes only represent changes to beliefs about the world state and its expected outcomes. The differences between them are not as clear in practice, a problem encountered by imitation research (Zentall, 2007). Associating an otherwise neutral event with a motivator gives that event motivational value, coupling the value of that event to its association. This means that while a meme may measurably alter the motivation to express another meme, the mechanism for the motivational change could remain unclear.

Persuasion and Attitude Change

Marketing and persuasion research examines the flip side of incentives, how to increase the attractiveness of attitudes and activities. Memes interact strongly with attitude change. Memes can be used as a tool in marketing, to spread an idea or behavior. Viral marketing concepts employ memes (Chielens & Heylighen, 2005). Alternatively, persuasion techniques can be used to alter the social environment for memes. Raising the demand for wine will drive more people to learn wine making, for example. Finally, a meme may be conceptualized as a message in a persuasion framework. Understanding memes as persuasive messages has interesting implications- a meme is a message capable of persuading the receiver to transmit it. This final linkage gives traction for understanding meme dynamics.

Persuasion research gives insight into the dynamics of meme motivation. For a single presentation, a meme can be directly considered as a persuasive message designed to inspire retransmission. For the considerable amount of time that may pass between reception and transmission, persuasion allows a mechanism for changing the meaning and value of a meme. Generic messages provide one manner to change the value of meme expression by altering related attitudes. Repeated presentations of the same meme are a special subset of generic messages. This discussion of persuasion will start with appeals of different persuasion processes, then note the different pathways to persuasion, and conclude with the implications for meme attitudes and expression.

Persuasive Appeals: The Tripartite Distinction

Current models of persuasion have been converging toward a tripartite model of persuasion (Wood, 2000). The Wood (2000) tripartite distinction considers three drives for attitude change: supporting the ego and identity, understanding the environment, and maintaining social relationships. These distinctions correspond with different classes of motivators noted in Section 3.4.3. Attitude change that supports a desired identity generates an intrinsic payoff. Understanding the environment helps an agent obtain extrinsic payoffs. Maintaining social relationships helps an agent elicit social rewards. Appendix D provides a representation and additional discussion of the motivators involved in attitude change under the tripartite view.

One method to affect the value of a meme relies on changing its implications for personal identity. This form of persuasion attempts to explicitly change the intrinsic value of behavior and ideas by associating them with a particular concept of identity (Wood, 2000). In order to improve motivation to express a meme, it may be framed as an expression of person's identity or a preferred identity that they wish to possess. In this context, meme expression defines who a person is and how they should feel afterward. Moral appeals and the opinions of role models employ this form of attitude change.

Two types of identity must be managed: self-identity and the identity impressed on others (Chaiken, GinerSorolla, & Chen, 1996). Self-identity connects with ego, as an person selects memes in order to express their personality. Managing impressions works differently, as a person may choose expressions that project a desired image. Since the actual opinions of other people are unknown, this identity still relates to the ego but is mediated by the inferred beliefs of others. In example, the difference between managing identity and impressions is one of the factors used to distinguish between guilt and shame (Elison, 2005). Note that neither desire needs to correlate with tangible consequences, even social ones. A person's preferred identities may be incompatible with their social rewards but still provide intrinsic value.

Motivation to express a meme can be altered by changing a person's view of the ground truth of the world. Rather than attempting to associate a meme with a construct such as popularity, this persuasion process changes context or implied consequences. Factual appeals succeed when they connect with a person's concerns for informational accuracy. New information provides a fundamental process for learning, making such attitude change normative (Petty & Cacioppo, 1986). Providing related information and appeals to reason follow this pattern. By changing the expected consequences, the motivation to express a meme can be altered. Changes to implied social rewards are an important subset of such consequences. These three types of appeals can influence an agent by multiple pathways, either by the direct content of a message or through the context.

Pathways to Persuasion: Dual Process Theories

Dual process theories highlight two different pathways for persuasive information, central and peripheral. The Elaboration Likelihood Model (ELM) theorizes a process by which this occurs (Petty & Cacioppo, 1986). While ELM is an older dual process model, it provides a good starting point for analysis as it captures the core insights of dual process persuasion. *Dual Process Theories in Social Psychology* by Chaiken and Trope (1999) provides updated and alternative models which may be useful in examining specific problems.

Dual process persuasion reconnects with the literature on filtering introduced in the discussion of attention in Section 3.4.1. Dual process theories of attention utilize similar concepts, filtering by semantics or by physical characteristics (Treisman & Gelade, 1980). Attention is a necessary but not sufficient condition for processing. A difference between levels of processing and levels of attention implies attended but unprocessed information. The distinction may depend on the time horizon in question, as attended and stored information might be processed later. For example, a very busy person may overhear a joke but only process it and laugh when their task is done. Experimentally differentiating unattended from unprocessed information may be intractable- the observed behavior will be identical.

ELM states that the central pathway for persuasion requires cognitive processing of information, while peripheral information is processed by low cognition heuristics (Petty & Cacioppo, 1986). The central pathway can consider the accuracy of information and its contingencies through deep processing. Central pathway information requires genuinely appealing semantics and implications. Peripheral information provides cues for opinion and action. While peripheral information may be pertinent, extraneous information also affects this channel. The halo effect is an example of a peripheral cue, allowing an attractive message source to boost the appeal of a message (Kelley, 1955).

In order for a meme to reproduce, a person must process sufficient information to replicate the meme. This attention constraint requires that central processing of meme signs must occur at least once before a meme *can* be reproduced. The signs of a meme must transmit through a medium and the medium is part of the larger environment. This means that contextual information surrounds a meme within the medium and environment. Contextual information will be processed within both the central and peripheral pathways. Contextual information and peripheral processing of meme signs have a major role in motivation, showing *why* a meme should be expressed.

Table 3.3: Pathway Effects on Meme Learning

	Content (Meme Signs)	Context (Source, Environment, Outcomes)
Central	Meme Semantics	Outcomes of Expression
Peripheral	Syntax, Exposures	Cues for Expression

Table 3.3 notes the type of information acquired by each pathway. The signs of a meme provide the core content. Central processing of meme signs allows learning of meme semantics. As noted earlier, re-expression can only occur when meme signs are centrally processed. During central processing, the core semantics of a meme will affect motivation. A chain letter stating "Send this to ten friends and you will have good luck" employs pure semantic persuasion. One segment of the memetics community studies the effects of semantic effects on replication, such as hooks and implied rewards (Bjarneskans, Grønnevik, & Sandberg, 1996, Fig. 2).

While the semantic information must be obtained to reproduce the meme, the semantics of a meme are constrained by their credibility. If a meme implies payoffs for expression, these payoffs might not be grounded in reality. A meme can contain lies, inaccuracies, and hidden assumptions. This is not necessarily a hindrance. Research shows that the persuasiveness of a message depends on its appeal, not its factualness (Wood, 2000). Bounded rationality makes this a necessity, since no human knows the true "factual" state of the world (Simon, 1982).

Credibility and discrepancy with existing beliefs act as moderators for the semantic appeal of a message. The counterpart of discrepancy, novelty, has been discussed previously in Section 3.4.1 on attentional salience. Findings discussed by Sternthal, Dholakia, and Leavitt (1978) note similar effects on message processing: low discrepancy messages suffer from low processing and high discrepancy messages are disregarded. High credibility counteracts the tendency to disregard high discrepancy messages (Sternthal et al., 1978). Perceived bias in sending a message reduces the fitness of the message, possibly due to a negative effect on credibility (Kelley, 1955).

Peripheral processing of meme signs works through an awareness of the message outside of parsing the semantics. Information with lower relevance will tend to be taken at face value, through heuristic cues (Oldmeadow, Platow, Foddy, & Anderson, 2003). Syntax and exposure effects act through peripheral cues. The syntax of a message provides peripheral cues that may alter its appeal (Howard, 1997; Sparks & Areni, 2008). Returning to the chain letter meme, introducing a spelling error could alter the fitness of the meme without changing the semantics. Exposure effects include the influence of independent sources and the repetition effect. Research shows that repeated exposures increase the cumulative attitude change from a message (Ray & Sawyer, 1971; Ray et al., 1971). Likewise, multiple independent sources of repetition have greater effect than a single source (Harkins & Petty, 1987; Schunk, 1987). The emergent phenomenon from this form of persuasion is known as the bandwagon effect (Leibenstein, 1950). Fads are a type of meme that rely on this type of persuasion.

The context consists of situational factors involved in meme expression. During meme expression, an agent interacts with the environment and some outcomes occur. The source agent, the environment, and the outcomes relate to the specific instance of meme expression. Every time an agent observes a meme expression, the appeal of expression will be affected by these contextual factors. The cumulative sum of exposures will determine the appeal.

The environmental context and perceived outcomes are connected. From a normative sense, an agent should be learning how meme expression in a certain environment relates to valued outcomes. This is an input-output relationship, as noted in the meme reproduction expression (Eqn. 2.1). The environment and outcomes may be either physical or social. Central pathway processing involves subjective estimation and valuation of the results of meme expression across different circumstances. The observed outcomes provide information about the payoffs of meme expression. Persuasion occurs because an agent observes desirable or undesirable outcomes from expression, as parameterized by environmental states and perceived capabilities of the source.

Peripheral processing of the environment and outcomes will generate cues for meme expression. Environments where meme expression occurs commonly can provide cues to express the meme in that setting, a form of occasion setting. Similarly, meme expression may become associated as a cue for certain observed outcomes purely as a result of repeated pairing. Environment and outcome cues can change the appeal of a meme by associating it with certain environments and results in a general sense. For both peripheral and central processing of these context effects, persuasion occurs through vicarious experience. According to Bandura (1986), the learning from observing a meme expressed by another person should be similar to that obtained by expressing the meme personally.

The source of meme expression provides a very different and important vector for attitude change about a meme. This portion of the context probably appeals to the intrinsic value of meme expression through social psychology factors. Central processing of the source's abilities provide information about the desirability and transferability of an observed behavior. The source increases the desirability of expression when they appear to have greater competence for deciding on behavior (Cialdini & Goldstein, 2004). A source's expertise can fit this mold. As the source's expertise in relevant decision making increases, the source's performed or prescribed behavior becomes more appealing (Cialdini & Goldstein, 2004).

Transferability relates to the ability of observed outcomes to generalize from the source to the receiver. By knowing the source, an observer can correct for their differences compared to the source when anticipating the outcomes of meme expression (Bandura, 1986). Source expertise in production ability might reduce a meme's fitness via this mechanism. For example, the average person will not feel competent to imitate the movements of a trained ballet dancer. Persuasiveness due to transferability interacts with self-efficacy and locus of control, since this affects attitudes about the control of outcomes (Schunk, 1987).

A second central processing persuasive factor is the connection between the source's meme expression and their self identity and social identity. Imitating agents with desirable social characteristics provides intrinsic motivation because it reflects desired identities. Role models and reference group members hold additional persuasive influence (Kameda, Ohtsubo, & Takezawa, 1997). If meme expression appears connected with maintaining a certain social identity, this will affect its motivators. Improved social relationships and status imply contingent social payoffs, making this at least partially extrinsic. Berger and Heath (2007) note that people actively diverge in behavior primarily to prevent misclassification in society. This divergence has been shown to be a function of the perceived dissimilarity of other individuals. Both paths may be likely in this case- people consciously and unconsciously value behaviors that project their desired social identity and avoid those which might cause them to be wrongly identified.

The peripheral cues of the source are at least as powerful as the central ones. Conformity, in-group bias, similarity, authority, and the halo effect alter the perceived value of imitation. These social biases exert a persuasive force which will be referred to as social influence. Social influence depends on an agent's intrinsic motivation to imitate other agents. In general, humans appear to have a drive to imitate other individuals, known as herding behavior. The Asch (1955) study showed that the level of social influence increases with the number of individuals performing a certain behavior. Tajfel (1982) qualified this result by establishing that members perceived to be in the same group (an in-group) have a higher influence than outgroup members. Perceived similarity can also be a persuasive factor, with similarity positively correlated with persuasiveness (Platow et al., 2005). Too much similarity has also been proposed as a reason to diverge however, which means this may form a bell shaped curve or an uncanny valley effect (Snyder & Fromkin, 1980). Authority can also exert a powerful influence, even without any coercive power (Milgram, 2004). The halo effect indicates that attractive and likable individuals have a higher influence (Kelley, 1955). These factors indicate that memes expressed by numerous likable members of the same group will tend to be attractive, even if the outcomes of expression appear negative. The effect of social biases as peripheral cues is not prevalent in meme-centric literature, but these effects have serious implications for the spread of memes. In particular, these source effects on meme indicates a strong influence related to the initial population expressing a meme.

ELM provides a good framework for examining the persuasive influence of memes, but it is not the only dual process theory. Alternative and updated dual process models may be found in Chaiken and Trope (1999). Examining the implications of other dual process theories may lead to additional insight into why certain agents are susceptible to certain memes. While ambiguities and controversies exist in motivation literature, motivation and persuasion cannot be sidestepped for serious discussion of memes. The interaction between motivation and meme expression will be used as the core of a model for meme replication expressed in Section 5.

3.4.4 Production

The production process of a meme involves a change in behavior based upon the semantics of the meme. Production involves expressing a meme, creating signs that might be observed by other individuals. Production can be a hurdle to replicating a meme, if certain populations are physically unable to express a meme. When production occurs, it introduces some variation in memes due to irregularities in the situation and muscle control. Production also offers two systematic opportunities for mutation in memes: differences in individual ability and multiple forms of expression.

Barriers to production may occur due to environmental factors. Certain behaviors interact with tools or other individuals. Such requirements constrain expression of memes. For example, participation in a three legged race requires a second person and sufficient space to run. The situational requirements for meme expression are interesting as they establish spots and contexts where memes spread. Identifying these contexts provides important information if one seeks to monitor or influence the spread of a meme. Spot oriented modeling seems suited to studying these barriers, such as those used in the Novani, Putro, and Deguchi (2007) pathogen model.

The barrier may be the result of different individual characteristics. These restrictions mirror the barriers noted for reception in Section 3.4.1. Expressing a meme may require certain physical capabilities that are not universal. A person without arms will be unable to imitate clapping, for example. These restrictions are specific to each meme expression. Within demographics, these physical differences may play a minor role. Children and adults of different developmental levels may have significant differences that affect production however. Bandura (1986) and Piaget, Tomlinson, and Tomlinson (1929) both imply that younger children may lack necessary faculties to imitate certain behaviors done by older individuals.

Differences in individual abilities do not always block meme production. Humans have a considerable ability to adjust for differences when replicating behavior. This returns to the concept of transferability noted previously in Section 3.4.3. A person may express a meme, but it is adjusted for their individual abilities. The difference may be trivial for observers, in which case the transmissions will be functionally equivalent. Some differences will result in different perceived semantics, however. These differences introduce variation and mutation, creating a variant of the original meme.

Returning to the concept of having no arms, Marty Ravellette was a man with no arms who drove a car and performed other tasks using his feet (Hayes, 2003). Reconsidering clapping, it is likely that Ravellette learned the cultural behavior of clapping from other members of society but produced it with his feet rather than his hands. This is a significant mutation. If Ravellette had been introduced into a society unfamiliar with clapping, foot clapping could spread as a meme. Such a population is not as uncommon as one might think- children are regularly learning cultural behaviors. While not all differences in abilities are this extreme, even small differences in capabilities can introduce systematic mutations in meaning through many replications.

A second type of variation for memes results from multiple different expressions of meme information. Bandura (1977) considered observed behavior, spoken descriptions, and written descriptions as three fundamentally different ways to transmit a behavior. A study by Zukow-Goldring and Arbib (2007) on affordance learning classifies these differently. Scientists taught affordances by giving scripted verbal instructions, interactive verbal instructions, guiding subjects' hands through the behavior, and performing the behavior for observation.

Examining these two sets of classifications, three factors seem to parameterize expressions: abstraction, interaction, and concurrency. Table 3.4 notes the types of expressions capable of transferring the meme information. Abstraction indicates if a meme is expressed directly or conveyed symbolically. Concrete replication means directly performing behavior related to a meme, such as using a technology or following a social norm. Symbolic replication would include a lingual description, such as instructions or a manual. Some memes do not have a concrete equivalent and all of their expressions are symbolic, such as a metaphor. Interaction refers to the ability of an observer to provide input into a meme expression. One example of interaction is a second agent asking for greater detail or a rephrasing of a concept. A second type of interaction relies upon a second agent to act in a particular way to complete a meme, such as a knock-knock joke. Memes that spread through direct instruction and engagement are interactive, while those learned by passive observation are non-interactive. While concurrency is left out from this table, each meme theoretically has a concurrent and non-

Table 3.4: Types of Meme Transmission

	Abstract/Symbolic	Concrete
Non-Interactive	Description	Observed Behavior
Interactive	Dialog	Collaboration / Assistance

concurrent form. Concurrency refers to if a meme is expressed at the same time that an agent attempts to emulate the expression. An instruction taught using "repeat after me" takes advantage of concurrency. Some types of instruction, such as guiding a child's hands, require concurrency. These types of meme expression noted in Table 3.4 have different implications with regard to the type of variation introduced. Concrete expression allows an observer to view the complete expression of a meme, situating the behavior in the real world. In theory, this provides the full information required for replication, but in practice an observer may not attend to the right details or be unable to sense important information. Additionally, variation from individual differences will be introduced. Since the transmission is direct, these errors in transmission will be primarily syntactic. Conversely, an abstract expression will tend to follow a code or language of symbols that reduces misunderstandings due to syntax. However, reducing a complex behavior into a symbolic notation allows for ambiguous statements and descriptions. Resolving these ambiguities could result in differences between the source's description and the receiver's understanding. For this reason, semantic variation appears more likely for symbolically expressed memes.

Interaction allows for communication protocols and other systems that reduce variation (Schramm, 1963). Requests for repetition provide a mechanism to reduce syntactic mutations, while requests for rephrasing can reduce semantic errors. Instruction and academics rely upon interaction to ensure that information is faithfully transmitted. If students were only given books or rote lectures, small errors could accumulate over generations and undermine the curriculum.

Production of memes plays an essential role in the evolution of memes. Variations can be introduced due to syntactic errors and semantic ambiguities. The production process for a meme, including any symbolic notational system, will regulate the extent and nature of meme mutation. Linguists and semiotic researchers study the mutations introduced within language and common symbolic systems (Christiansen & Kirby, 2003). For concrete expression, the errors introduced may be specific to an implementation. While the later section on model implementation does not focus on production processes, this is a sacrifice made to focus the model on a particular set of parameters. Production processes warrant study and hopefully memes will be explored further within this subdomain.

3.5 Conceptual Model

Pulling together the theory from the prior sections, a full model for memes comes together. This conceptual model uses the synthesis of observational learning (Bandura, 1986) and information theory (Shannon, 1948) to organize the mechanisms that guide meme evolution. This conceptual model has two parts: the cognition and the transmission. The cognitive models are consistent across situations, because they relate to human thought processes. While they are parameterized by individual differences and learning, the underlying psychological mechanisms have been found to be consistent across individuals and experiments. The model for transmission is different, varying as a function of the message, the noise, and the medium.

Figure 3.10: Meme Conceptual Model

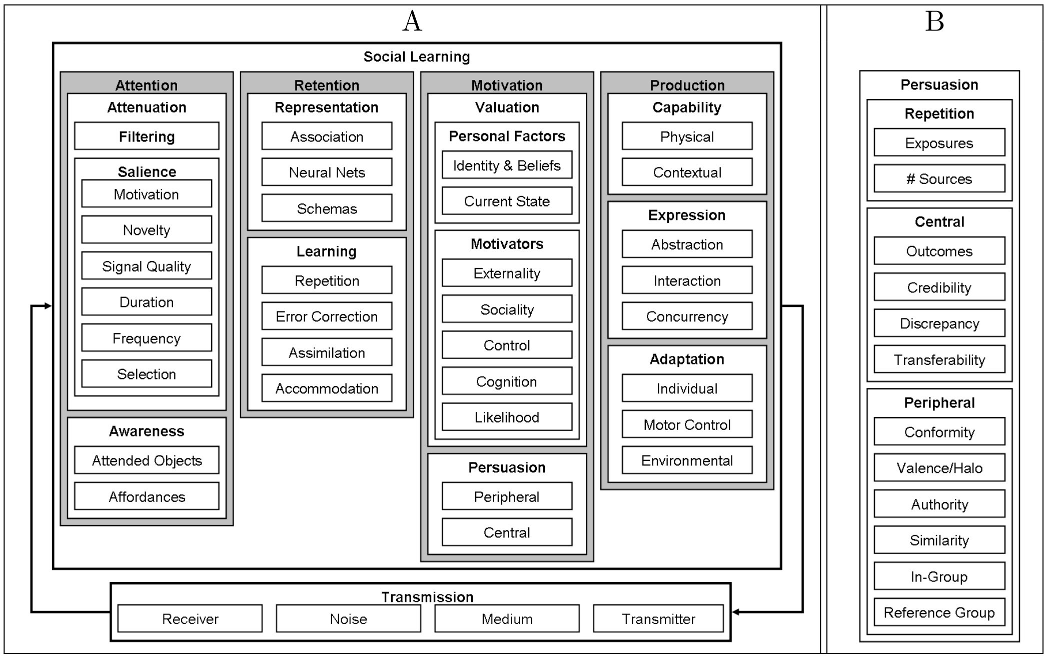


Figure 3.10 shows the layout of theoretical concepts as they related to memes. Part B displays the persuasion conceptual model, which is too detailed to be contained with the main figure. Each of the concepts mentioned maps to one or more theories from social science mentioned in the previous section, including the container concepts. The implementation described in Chapter 5 is derived from this conceptual map and its related theories, connecting them together

into a workable model. The theories pertaining to retention and production are implemented only in a rudimentary fashion, since their dynamics are not vital to the scenarios examined. However, they are still important parts of the conceptual model for memes that deserve further study.

This systems approach to modeling memes provides a useful way to explain reproduction, variation, and selection. However, this model cannot be useful unless it can be applied to real world problems. To usefully apply the model to memes, there must be a way to connect memes to empirically measurable properties. The following section addresses this, focusing on the observability of memes.

Chapter 4

Observability of Memes

The downside of examining memes as a form of semantic information is that learning this information is not necessarily observable. To definitively prove a meme exists, the process of recursive reproduction described in 2.4 must be measured. Transmission is not the only process which can be examined for memes, but it is the most fundamental for measurement. If memes cannot be identified and measured, it is not possible to measure their mutation or their competition.

Once this process is well understood for a meme, it is straightforward to study competition between memes by measuring more than one at the same time. Studying variation requires identifying memes and tracing their variants, which also requires an understanding of the transmission processes. For the ultimate goal of measuring the evolution of memes, it is necessary to measure all of these processes. However, since transmission is not yet well understood, it would be premature to suggest analytical methods for these more complicated processes.

This section will start by discussing different approaches to measuring and observing memes. Ultimately, all measurement of memes depends on measuring behavior or changes in behavior. After this is established, a particular type of meme that is amenable to measurement will be described. This meme, known as a socially learned affordance, is information about a possible action in the environment. This meme, unlike many memes, has a direct behavioral expression that can be measured.

4.1 Measuring Meme Transmission

Memes cannot be useful within scientific discipline without methods of measurement. Prevalence and transmission provide useful information for the study of memes. Prevalence of meme-related behaviors can provide a metric for the prevalence of a meme through positive or negative correlations. Meme transmission gives the ability to discern the existence of a meme from other types of learning and behavioral change. Meme transmission can be examined as a pathology and is sometimes simulated using vector-host models (Aunger, 2002).

Table 4.1: Meme Transmission Processes for Measurement

Behavior	**Reproduction**	
	Reproducing	Equilibrium
Activated	Diffusion of Behavior (Adoption, Innovation)	Entrenched Behavior (Normal Response)
Inhibited	Displacement of Behavior (Abandonment, Closeting)	Entrenched Aversion (Taboo, More)

Two traits affect the applicability of these metrics: dynamics of reproduction and behavioral activation. Table 4.1 notes categories based upon these parameters. The reproduction dynamics indicate whether a meme is currently reaching new people or if all agents in the sample already know the meme. The behavioral activation indicates if the meme works by increasing particular behaviors, as opposed to only inhibiting particular behaviors.

For any given population, a meme can either be reproducing or at equilibrium. If a meme is reproducing, some individuals have the opportunity to spread the meme to some other individuals who could also spread the meme. Equilibrium occurs when all possible receptive individuals already possess a meme, preventing reproduction. A carrier of a meme may update the information related to a meme when a new meme is presented. This process will not be considered reproduction, but still indicates meme activity. Saturation is a special case of equilibrium where all possible agents have received a meme. When a meme saturates the population, reproduction becomes impossible. Fully entrenched memes cannot reproduce.

Measurement of memes also depends on how they activate behavior. Every meme must create some change in agent behavior in order to reproduce. Behavioral activation can be excitatory or inhibitory. An excitatory meme increases the use of a particular behavior, such as a catch-phrase. If a meme is associated with a unique behavior, that behavior might be considered the behavioral expression of the meme. On the converse, an inhibitory meme spreads through the conspicuous absence of certain interactions. Memes can also use both mechanisms, exciting some behaviors and inhibiting others.

4.1.1 Measuring Diffusion of Behavior

Memes are most amenable to analysis when they are reproducing. Reproducing memes actively spread or die out in a population. Changed behavior and awareness of new behaviors provide reasonable metrics for meme reproduction. Transmission cannot be definitively measured unless reproduction occurs. When diffusion of a meme through a population causes individuals to do certain behaviors more often, this gives the best opportunity for measurement.

Firstly, this condition allows a researcher to observe reproduction of the meme. Such a learning measure requires two components. The first measure should infer if a person does not know the meme without teaching the person the meme. Using sequential measurements of this sort, a researcher can determine when an agent learned the meme. The second component is measuring expression of the meme. Meme expression would be the behavioral patterns increased by knowing the meme. Such expressions are the opportunities for other agents to learn the meme. By measuring expressions, a researcher can determine when a person starts transmitting the meme for others to learn. If this process is measured from start to finish, the existence of a particular meme can be empirically verified. Additionally, this approach helps define the semantic information of the meme since learning is measured separately from expression.

If learning cannot be directly measured, examining memes is more complicated. Direct measurement of learning may not be possible due to the possibility of learning the meme during the process. One cannot simply ask, “Did you hear about the new iPod Nano sale?” since this question causes the person to learn about the sale. Since indirect measurement is necessary, it may be unreliable or cost-prohibitive for a large-scale study.

As an alternative to measuring learning, a measurement of diffusion can be examined by measuring active expression of memes alone. By measuring when agents are exposed to a meme and when they express the associated behavior, causality can be inferred between exposure to the meme and the first time an agent expresses a meme. If an agent is unlikely to perform a behavior for their first time until they are exposed to a behavior, this indicates that a meme is present. In this way, the time of an agent’s first expression of a meme can be a metric for learning. However, since this approach does not directly measure learning it cannot definitively prove that a meme exists. This measurement approach only shows that people act as-if a meme exists. Also, it gives far less information about the exact meme involved. Since the same behavior may transmit different information due to the context, this approach gives less information about what has been learned, if anything.

Simple measurement may also be impossible since a meme may increase the distribution of certain behaviors, such as the frequency that they occur, rather than the specific behaviors that occur. Such memes will be more complicated to study, since statistical tests must be applied to infer that behavior has increased by a statistically significant margin.

Even with these caveats, measuring memes reproducing by activating new behaviors or patterns of behavior is the most straightforward case. This case allows measurement to show that reproduction occurs and allows insight into what information is spread by the change in behavior. As such, these memes give the best opportunities for studying meme transmission and prevalence.

4.1.2 Measuring Displacement of Behavior

Measurement of such memes can be measured by the expansion (diffusion) and contraction (displacement) of behavior (Heylighen, 1998). Displacement is used in place of the more common term, abandonment, because a meme may force a behavior to be practiced in secret. For memes, this is functionally equivalent because a person's social behavior would imitate abandonment.

Measuring memes through displacement is slightly more complicated than measuring them through diffusion. When learning a meme increases a certain behavior, new carriers will tend to take the behavior. If learning a meme tends to suppress or extinguish a behavior, a researcher must look for a decrease in that behavior or examine the last time that a person uses such behavior. In this way, it is straightforward to examine persons who have probably learned the meme. They previously used the behavior regularly, but after learning the meme use it less.

The additional difficulty comes from measuring exposure to the meme. If the meme inhibits a behavior, there are no distinct new behaviors to observe. Instead, reproduction requires that a behavior is expected to occur but did not. This means that a researcher has to know the contexts where the inhibited behavior typically occurred, prior to a person learning a meme. This means that a researcher needs a third measure: one to determine the contexts where a person takes the inhibited behavior. If these contexts are unknown, it is impossible to tell if an agent's use of a behavior has been inhibited. It is only possible to measure inhibition if the previous base rates were known.

If this difficulty can be overcome, the reproduction of inhibitory memes can be studied very similarly to the reproduction of memes that increase certain behaviors. In both cases, reproduction of memes means that it should be possible to prove that a meme exists and to examine its spread through society.

4.1.3 Measuring Entrenched Behavior

Once a meme has become universal, if measurable behaviors exist then a meme should be considered to be active at equilibrium. Memes of this nature will continue to "preach to the choir." Equilibrium occurs when all possible receptive individuals already possess a meme, preventing reproduction. A carrier of a meme may update the information related to a meme when a new meme is presented. This process will not be considered reproduction, but still indicates meme activity. Saturation is a special case of equilibrium where all possible agents have received a meme. When a meme saturates the population, reproduction becomes impossible. Fully entrenched memes cannot reproduce.

Entrenched memes are a tougher target for analysis than reproducing ones. Without being able to observe the transmission of information, it is difficult to determine its origins. This makes an entrenched meme difficult to distinguish from a reaction to shared environment or genetic factors. Cross-cultural and historical analysis methods from anthropology provide some insight into these problems but are not useful when the observer's perspective misses important information (Ruby, 1982). Assuming the existence of a meme can shed light on a blind spot but runs the risk of forming false hypotheses.

Testing for the existence of a meme can be important contribution to science, however. Giddens (1986) theorizes that a significant portion of human behavior is constrained by structural characteristics of the society, where unconscious assumptions about the context drive behavior. In some cases, these rules may benefit society by improving the outcomes of unintended consequences. In other cases, these blind spots may be the result of path dependent effects, rather than true reflections of predisposing factors (Margolis & Liebowitz, 1995).

Blind spots tend to be supported by "just-so" hypotheses, which can be exposed by identifying the mechanisms for adoption. Bans on women in the US army historically followed an entrenched meme of this nature. Females were deemed unfit as soldiers because women historically had not been soldiers. This ban began to break down in WWII with the development of the Women's Army Corps and official restrictions on women in combat were lifted in 1994 following the Gulf War. Role models and counterexamples break down cultural blind spots and inhibitory memes, a key element of social movements (Brown et al., 2004). Pluralistic ignorance is an example of a meme which can be displaced in this way (Prentice & Miller, 1996). During this process, the decline of a meme can be measured as a counter-meme. In this way, memes can be thought of as one mechanism fitting into the process of structuration (Giddens, 1986).

At equilibrium, new behavior cannot easily be measured. In active equilibrium, memes can be measured through behavioral expressions that would potentially spread a meme. However, widespread expression of behaviors may have alternative explanations such as individual learning from non-social cues. The problem is similar to measuring the potential for flu contagion in a epidemic by counting the number of coughs. Coughs will spread the disease to new hosts, provided a flu virus causes the coughing. Without proving that a current host can spread the coughing to a new non-coughing host, no proof exists for the existence of the virus. The excessive coughing could be the result of air quality or genetic factors within a society.

Aggregate measures of behavior may be useful for examining meme prevalence but cannot prove the existence of a meme. In practice, a meme can never be in full equilibrium however. Since new infants are blank slates with respect to memes, children are always entering society who are unaware of particular memes. As such, it should be theoretically possible examine children to determine if a behavior emerges from learning a meme or if it is simply a result of individual learning within a shared environment. Appendix B notes some of the ways that imitation can be distinguished from other mechanisms that alter behavior.

Measurement of reproduction over a limited and vulnerable sample, such as children, may not always be practical. As an alternative, it may be worthwhile to search for meme variation. An entrenched meme may have multiple variants, under a more fine-grained level of analysis. As such, in some cases it may be possible to break the analysis of an entrenched meme into a problem of examining a set of similar memes that actively reproduce. However, this requires examining the differences in learning and expression that are specific to each particular variant. This reduces to an issue of speciation, explained in more detail in Appendix A.

In general, limited analysis can be performed of memes that are at equilibrium. Only the prevalence of meme expression can be reliably measured. As such, the preferred solution to examining memes in this case would be to attempt to re-formulate the analysis to look at population subsets where the meme is not at equilibrium or to change the level of analysis of the meme to look at sub-variants that are not at equilibrium.

4.1.4 Measuring Entrenched Aversions

Memes can have an active equilibrium or passive equilibrium. Active equilibriums display behavior excited by memes, as described in the prior section. Memes at passive equilibrium represent inhibitors or have conditions for exciting behavior that are contingent on encountering a non-carrier. For example, a meme against smoking could inhibit smoking and include a verbal reaction when in the context of a smoker. Measuring memes at passive equilibrium is the hardest case.

Memes in passive equilibrium are harder to measure as they represent inhibitors. The prevalence of inhibitory memes may only be possible to measure through perturbation- aversion against certain behaviors. An excitatory meme spreads a behavior which would otherwise be less likely to perform by chance, since then the behavior will be learned socially rather than independently. Conversely, an inhibitory meme may not spread unless the associated behavior can be readily deduced- the person needs to know that the behavior exists before they can inhibit it.

Measuring inhibitory memes at equilibrium may only be possible through cross-cultural comparison. Western practices inhibit eating spaghetti raw. Left to discover spaghetti when hungry, a person might easily begin to start chewing on it. For a person encountering spaghetti within Western culture, the conspicuous absence of people eating spaghetti raw can spread such a meme. The new person will avoid the spaghetti and can similarly propagate the meme as a result. For a less obviously edible food such as cacti, inhibition of behavior might not be sufficient to propagate the information.

Children can initially fail to understand inhibitory memes. Martin Luther King Jr. related a powerful example of this in his autobiography:

> The climax came when he told me one day that his father had demanded that he would play with me no more. I never will forget what a great shock this was to me. I immediately asked my parents about the motive behind such a statement (King, 1998).

Likewise, child abuse victims often take many years to realize that their experiences violate cultural norms (CrossonTower, 1999). While this paper focuses on memes that may be actively measured by reproduction or activity, analyzing and manipulating inhibitory memes has broad implications for social justice.

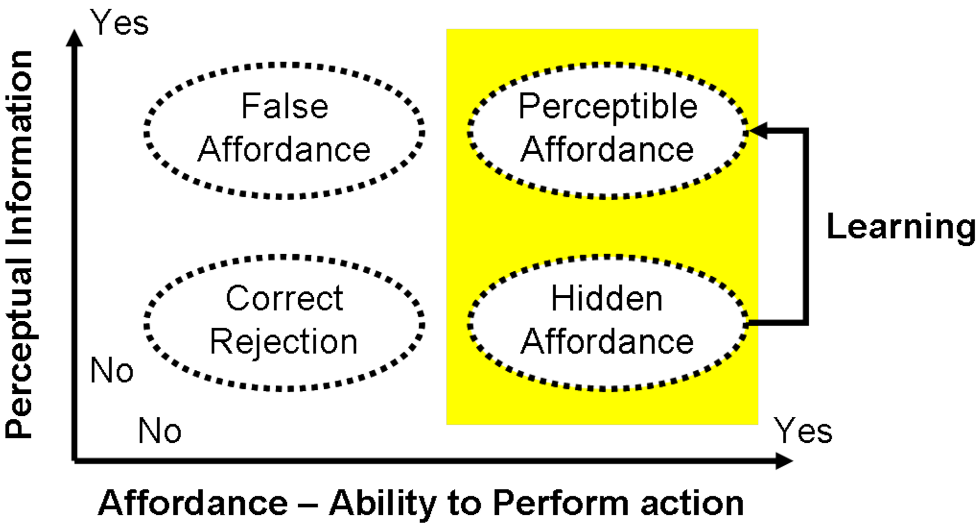
Measuring a fully entrenched inhibitory meme may be impossible in some cases. With no obvious signs to indicate that expression occurs and no clear indications that learning occurs, there is very little to measure. Inhibitory memes cannot be measured for prevalence, though their associated behaviors may be examined for relative prevalence between different cultures. However, inhibitory memes may still be amenable to study in smaller populations by using cross-cultural comparisons or learning among children.

4.2 Socially Learned Affordances

One type of meme that is particularly amenable to measurement is a socially learned affordance. In perception theory, an affordance is a relationship between an organism and a part of its environment that allows a particular type of action (J. J. Gibson, 1979). For example, a human has the affordance to swing a hammer. A goldfish does not have this affordance, as it has no hands. The ecological approach to perception posits that the environment is perceived in terms of the affordances that it offers, referred to as direct perception. Affordances always exist- they represent the potential for action.

Affordances are not always known, however. For example, a hidden light switch always offers the affordance to be turned on by pressing it. However, until the switch is identified it represents a “hidden affordance.” A hidden affordance is a potential for action that an organism is not aware of yet. Likewise, a fake switch may not offer an affordance but could be misidentified, a "false affordance." As shown in Figure 4.1, Gaver (1991) framed this issue using two orthogonal aspects: 1. Is an affordance available? and 2. Is the affordance perceptible?. By learning an affordance, an agent moves from having a hidden affordance to having a perceptible affordance (known affordance). In this way, an agent becomes aware of a new action opportunity. Social learning is one way for this learning to occur.

Figure 4.1: Relationship Between Affordances and Perception. Adapted from Gaver (1991)



Social learning is important because the space of possible actions for human interaction is vast. Even easily inferred actions can remain unknown, simply due to competition for attention. Learning by observation greatly reduces this space, exposing an agent to the affordance in practice. Observing such an action can allow an agent to discover an affordance. If affordances are inferred by direct observation, this corresponds to learning by imitation. Alternatively, affordances can be learned indirectly through verbal descriptions and other social mechanisms.

This behavior fits the basic requirement for a meme: it is information that replicates socially. Affordances also fulfill the requirements for evolution. Reproduction can occur socially, as by imitation. Variation is introduced when performing and observing the action. Competition occurs between observing affordances and attending to other environmental information.

The possibilities to open a door, buy a product, or clap hands are all affordances. Affordances have varying degrees of variation. An affordance such as using a doorknob is unlikely to have much persistent variation. Conversely, imitating a dance sequence will be prone to significant variation. As a result, affordance learning gives the opportunity to study evolving and non-evolving memes.

Affordances are amenable for study because they are readily observable. Since affordance learning allows an agent to recognize new opportunities for action, it will typically activate behavior. Since agent learns about the behavior itself, this grounds the learning in a directly observable phenomenon. As such, a socially learned affordance can be used as an observable meme. Novel actions, such as learning a new computer GUI, or strategies, such as learning a new chess maneuver, can be examined as part of this category.

The semantic information of an affordance is observable- by intentionally performing the action, an agent demonstrates their awareness of the affordance. Affordance learning requires that an agent becomes aware of a new action possibility. This possibility may be a specific action available on a specific object in its environment, or a more general learning about the possibility to perform an action when certain conditions exist in its environment. Either way, such learning is a necessary requisite for the agent to intentionally perform the action. When agent performs an action, they demonstrate that they have learned about that affordance.

However, this does not prove that the affordance was learned socially. This is because we cannot directly measure learning or knowledge. This is a common issue with learning research- even if learning appears to have occurred, multiple kinds of learning could be responsible for the same behavioral changes. Even in the most direct form, imitation, researchers must take great pains to ensure that behavioral changes are caused by learning an action rather than mimicry or by increasing the attractiveness of an already-known action Zentall (2007).

Even if one can prove the diffusion of a behavior- this does not prove the learning of the affordance for that behavior. Even if the existence of a meme can be demonstrated, where recursive social learning leads to the diffusion of an action for the first time, this does not guarantee that any affordance was socially learned. For example, imagine a species who knows how to eat berries (an affordance) but naturally fears that red berries are poisonous. If an innovator ate red berries and had no negative outcomes, other members of the species might observe and lose their fear of berries. This would cause a diffusion of behavior, without a diffusion of an affordance. The meme in this case would be the knowledge that red berries are non-poisonous. For this reason, a field study will at best be able to use diffusion of behavior to indicate that some meme exists- not be able to prove that the affordance was the meme.

Despite this limitation, treating the spread of behavior as-if it were the spread of affordance awareness is a useful approach. This approach can demonstrate that a meme appears to exist, where that meme is either awareness of the affordance or some piece of information that significantly increases the salience or attractiveness affordance. If the study population appears unlikely to have prior awareness of the action (i.e. they never performed that action before), this implies that affordance learning is the most likely mechanism. In this case, assuming behavior spread due to affordance learning may be a reasonable assumption. Moving from reasonable assumptions to proving affordance learning requires deeper measurements. To measure the type of semantic information transmitted requires complementary measurement approaches that help rule out other types of learning that increase the salience or attractiveness of the affordance. With that said, in many cases a reasonable assumption may be sufficient for the problem at hand.

Socially learned affordances offer a straightforward way to examine meme reproduction. If a subset of a population is unaware of an action but can learn it by observing the action being committed, then this is sufficient for a basic model of meme transmission. Models of technology diffusion have employed this principle (Windrum, 1999). Diffusion of innovation may be considered a form of affordance learning, since new adopters must first become aware of the ability to use a product. However, the usefulness of the model depends greatly on its mechanisms. When the discovery process is handled by a simple model (such as a sigmoid equation), the mechanisms may capture only rates of learning. A diffusion model using cognitive agents does not just tell how fast diffusion occurs, it gives insight into who will learn it and why. In the next section, an agent-based model for simulating socially learned affordances is presented which utilizes key cognitive mechanisms described in Chapter 3.

Chapter 5

Model Implementation: Agent Based Approach

Memes reproduce as part of a complex adaptive system. A complex adaptive system consists of subsystems and relationships, capable of creating emergent behavior (Holland, 1998). Agent based simulation is an effective method for examining emergence in this situation. It enables simulation, analysis, and validation beyond what is possible in a classical mathematical analysis. This allows simulating a variety of circumstances and dynamics that do not necessarily have closed form solutions. A second advantage is correspondence: cognitive agents are an analog for the meme system, allowing implementation of theory that can be explained and revised.

While the theoretical exploration examined meme evolution in general, the computational implementation focuses on a limited category of memes. Firstly, this implementation is tailored to model transmission of a specific type of meme, a socially transmitted affordance. While this is a subset of memes, it provides a starting point for studying meme reproduction- the fundamental mechanism of memes. Secondly, this implementation assumes high fidelity copying of socially learned affordances (no variation), without copying errors or transmission errors. Meme variation mechanisms were not implemented for two reasons. Firstly, the processes that drive variation of socially learned affordances have not been studied extensively- more empirical study is required to model these descriptively. Secondly, variation and mutation of memes were not important processes for the scenarios modeled using the computational model, as described in Section 6. This means that the implemented model is incapable of representing full meme evolution, and that many of the theoretical contributions from Shannon (1948) Information Theory are underutilized by this implementation. As such, it should be stressed that this computational model is a limited form of the larger conceptual model for memes described in Chapter 3.

The implementation of the model is an agent based simulation, a subset of computational models. Computational models are themselves a subset of math models, enforcing explicit representation of a model. The key advantages of a computational model are explicit representation of data, simulation, and testing against empirically collected data sets. For standard pen and paper mathematical model, these methods are extremely limited since they require explicit derivation. For open form or chaotic models, computational simulation may be the only way to examine a non-trivial form of the problem. The main disadvantage of a computational simulation is that a computer simulation must ultimately be discrete and Markov in its implementation.

The system for meme reproduction is an environment containing multiple agents, each capable of behavior and observation. This setup requires models for the environment and agent cognition. The environment, depending on its complexity, may require multiple models for interaction and dynamics. Human cognition involves submodels such as emotion, attention, stress, and decision making. Models of cognition may make use of many constituent models, such as decision making strategies that rely on selecting between heuristic strategies (Gigerenzer & Todd, 1999). However, design practices must be followed to prevent models from losing their connection with theory.

Table 5.1: Mappings from Conceptual Model to Computational Model

Correspondence	**Time**	
	Instantaneous	*Dynamic*
Single Module	Direct	Metric
Many Modules	Composite	Emergent

These practices are explored in depth in Appendix E. This explains the rationale for a model of models approach, mapping conceptual models to computational models, and agent based modeling. A primary take away from this tangent is the types of mappings of literature concepts to computational models. Table 5.1 notes types of implementations possible. A direct implementation realizes a conceptual model as a cohesive module in code. A composite implementation realizes a conceptual model from parts of multiple modules. While it may not be possible to directly correlate any cohesive piece of code to a composite implementation, all mechanisms and data for the model exist at each point in time. A metric implementation represents a conceptual model through its dynamics over time, such as a how a cellular automata can represent market equilibria. An emergent implementation does not exist at any given point in time, but is evident in the dynamics of the computational model over time. Each implemented model fit one of these forms, with the model for meme transmission being an emergent model out of these.

5.1 Agent Based Simulation

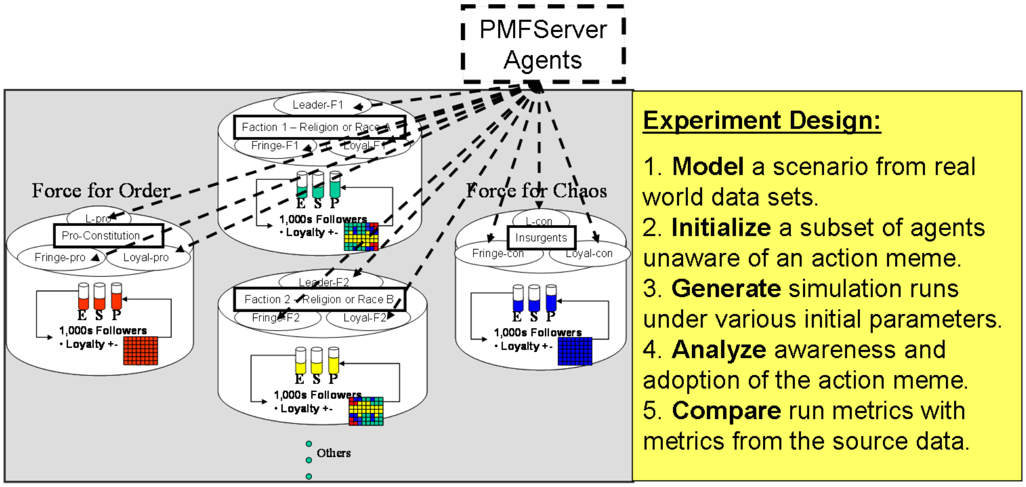
Memes will be studied by implementing an agent based simulation where socially learned affordances can be studied as memes. Socially learned affordances were chosen due to their observability, as noted in Section 4.2. Building an agent based simulation requires three basic designs: the agent, the scenario, and the simulation. The agents for this simulation are cognitive agents, requiring extensive design work. This is very different than the agents used for a typical cellular automata, such as Miller and Page (2004). The scenario is the full environment, consisting of the arrangement of agents and other entities. The scenario contains all the information necessary for simulation at any one time, including the mechanisms for interaction between entities. The simulation handles sequencing and updating of the scenario to move through time.

The agent based framework used for simulation is PMFServ. PMFServ is a modeling and simulation framework incorporating cognitive agents through a model of models approach. A PMFServ standard agent has models for perception, physiology, stress, emotion, personality, decision making, and basic social psychology (Silverman, Johns, Cornwell, & O'Brien, 2006). These models are based on respected social science theory such as the Janis and Mann (1977) coping style model. Variants of the standard agents have been used for crowd simulation (Cornwell, Silverman, O'Brien, & Johns, 2002), factional group simulation (Silverman, Bharathy, Nye, & Smith, 2008), and strategic leader simulation (Silverman et al., 2007). Agents in PMFServ can be extended by adding additional models and connecting them with existing models.

The PMFServ framework also allows for non-agent entities such as objects and groups. Objects are inanimate, but may be perceived and acted upon by agents. Groups allow explicit representation of membership and authority structures, which can be important simulations accommodating social identity. The FactionSim model family uses these capabilities, which are explained in detail in (Silverman et al., 2008). Figure 5.1 shows the FactionSim design, accompanied by the modeling methodology. FactionSim treats groups using a containership pattern, where an agent is either in a group or out of a group. It is an external social identity, as opposed to an agent's internal preference.

A PMFServ scenario contains a collection of agents, objects, and groups. These objects are designed to represent the environment in PMFServ, allowing opportunities for interaction. The simulation is used to coordinate their interactions over time. These will each be discussed in depth.

Figure 5.1: FactionSim Diagram and Methodology



Experiment Design:

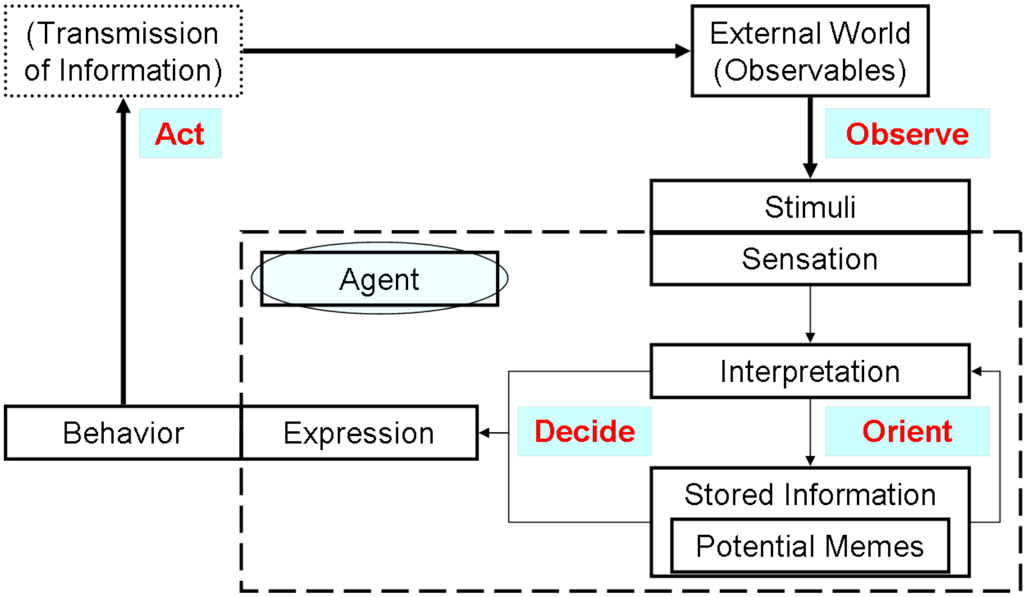
1. **Model** a scenario from real world data sets.
2. **Initialize** a subset of agents unaware of an action meme.
3. **Generate** simulation runs under various initial parameters.
4. **Analyze** awareness and adoption of the action meme.
5. **Compare** run metrics with metrics from the source data.

5.2 Cognitive Model Architecture

The Bandura (1986) model of observational learning was the central theory for the conceptual model for cognition. This can be mapped to a computational model for agents, known as the OODA loop. An OODA loop agent follows a stimulus-behavior pattern whose steps include observing, orienting, deciding, and acting (Tweedale et al., 2007). Revising Figure 3.5, the system may instead be expressed as in Figure 5.2. OODA loops have received significant attention in studying individual and organizational behavior, particularly decision making (Tweedale et al., 2007). The block diagram deviates from a classical OODA loop terminology, which may make mapping these two systems ambiguous. The OODA steps are noted with grey labels within Figure 5.2. Memes in an OODA loop framework rely upon an agent acting out a meme. Other agents may observe the meme. Some of these agents may orient to this meme through a learning process. Given a certain orientation to events, an OODA agent will decide to act such that the meme will be transmitted once more. OODA agents are of interest for simulation, since software practices and frameworks exist for implementing OODA loop agents. PMFServ implements OODA loop agents, which can be extended to fulfill the requirements for meme reproduction.

The PMFServ agent has cognitive models which handle perception of events and decision making. It also has an advanced subjective utility function, based upon a hierarchy of importance weights known as a GSP tree. The GSP, short for Goals, Standards, and Preferences, is the personality of an agent. Each tree is a hierarchal set of nodes, where the weight assigned to each node determines its relative importance for decision making. In theoretical terms, this tree may

Figure 5.2: Agent OODA Loop

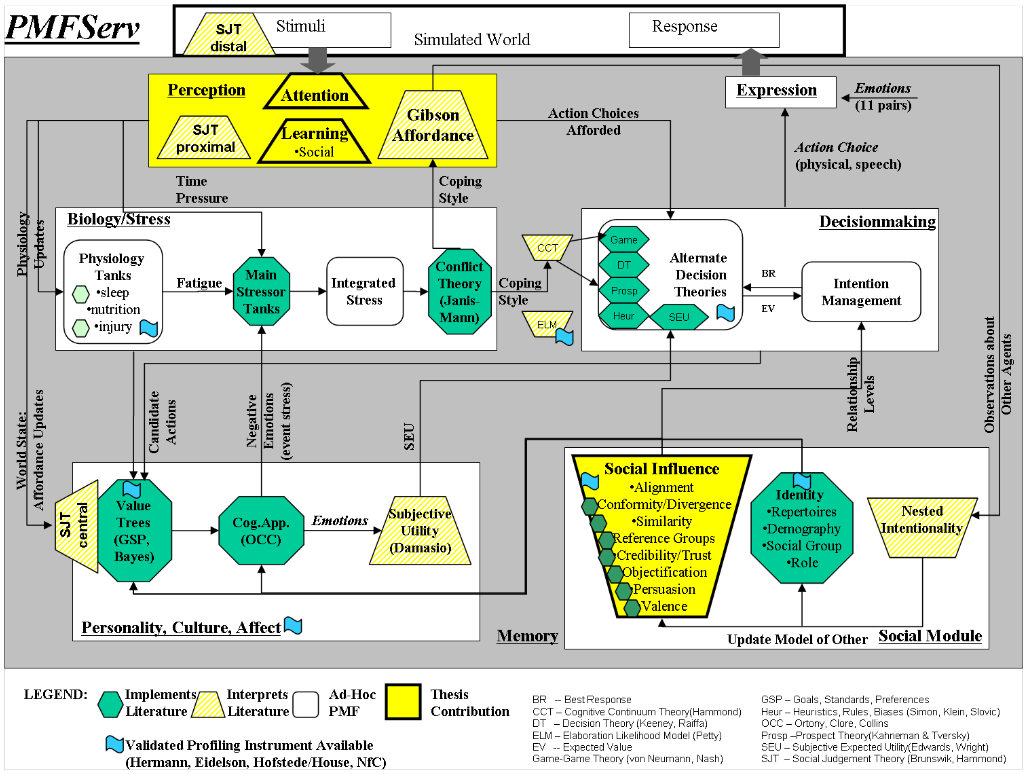


be considered to use Bayesian or importance weights- depending on the usage. Differences in the structure and weights of this tree change an agent's valuation of actions and outcomes, explained in greater detail in Silverman et al. (2006). A standardized tree structure is provided, which is based upon factors based partly on trait theory from Hermann (2003), House (2004), and other researchers. The PMFServ framework marks up actions with their motivators, referred to as "activations" for behavior. The tree structure is flexible enough to implement biological, cognitive, material, and social motivators. Valuation of actions is a function of the GSP and motivators for an action. In PMFServ, these motivators are referred to as "activations" for behavior and may be positive or negative. Each activation corresponds to a GSP node. The functionality provided by the GSP tree and subjective utility are sufficient to handle the valuation necessary for the motivation step of observational learning.

The standard agent in PMFServ has been made capable of spreading memes by adding new models and extending existing models. Meme processing required new cognitive models for attention and learning. It also required extensions to the perception and social modules. Figure 5.3 shows a diagram of the PMFServ module structure, with the extensions required for memes highlighted. The attention, perception, and social influence model implementations are intended to be general enough to work for a variety of different types of memes.

The connections between these models are noted in Figure 5.4. The major additions will be discussed in the following subsections on attention and social

Figure 5.3: Meme-Capable PMFServ Agent


PMFServ
SJT distal
Stimuli
Simulated World
Response
Perception
Attention
SJT proximal
Learning •Social
Gibson Affordance
Action Choices Afforded
Expression
Emotions (11 pairs)
Action Choice (physical, speech)
Time Pressure
Coping Style
Physiology Updates
Biology/Stress
Physiology Tanks •sleep •nutrition •injury
Fatigue
Main Stressor Tanks
Integrated Stress
Conflict Theory (Janis-Mann)
Coping Style
CCT
ELM
Decisionmaking
Game
DT
Prosp
Heur
SEU
Alternate Decision Theories
BR
EV
Intention Management
World State: Affordance Updates
Candidate Actions
Negative Emotions (event stress)
SEU
Relationship Levels
Observations about Other Agents
SJT central
Value Trees (GSP, Bayes)
Cog.App. (OCC)
Emotions
Subjective Utility (Damasio)
Personality, Culture, Affect
Memory
Social Influence •Alignment •Conformity/Divergence •Similarity •Reference Groups •Credibility/Trust •Objectification •Persuasion •Valence
Identity •Repertoires •Demography •Social Group •Role
Nested Intentionality
Update Model of Other
Social Module
LEGEND:
Implements Literature
Interprets Literature
Ad-Hoc PMF
Thesis Contribution
Validated Profiling Instrument Available (Hermann, Eidelson, Hofstede/House, NfC)
BR -- Best Response
CCT – Cognitive Continuum Theory(Hammond)
DT – Decision Theory (Keeney, Raiffa)
ELM – Elaboration Likelihood Model (Petty)
EV -- Expected Value
Game-Game Theory (von Neumann, Nash)
GSP – Goals, Standards, Preferences
Heur – Heuristics, Rules, Biases (Simon, Klein, Slovic)
OCC – Ortony, Clore, Collins
Prosp –Prospect Theory(Kahneman & Tversky)
SEU – Subjective Expected Utility(Edwards, Wright)
SJT – Social Judgement Theory (Brunswik, Hammond)

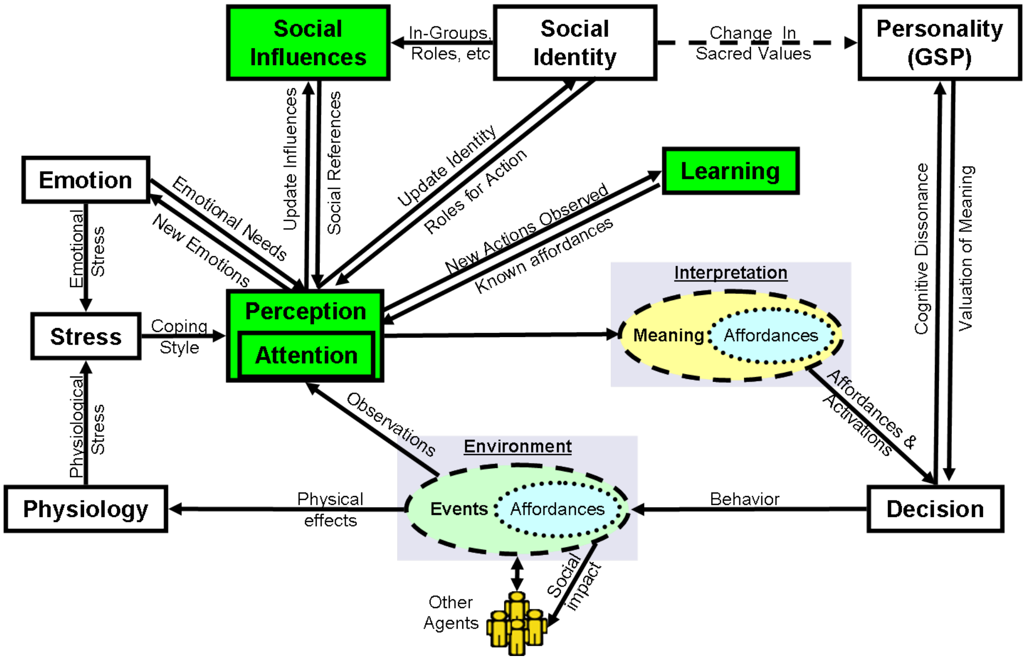
influence. It should be noted that for most of these theories, quantitative curves have not been derived. Most of these social science models state two possible conditions, with different behavior, or a curve which lacks exact definition or scaling. For this implementation, unspecified functional relationships for conceptual models are implemented using the simplest polynomial possible. However, the model design ensures that these functions are easily substituted for more complicated relationships.

Not all aspects of the conceptual model have been added to this implementation. A conscious decision was made to focus on the attention and motivational aspects of meme transmission, as opposed to the memory or production aspects. Figure 5.5 designates aspects of the conceptual model that have been added to the PMFServ cognitive model in order to model meme transmission.

5.2.1 From Event Processing To Social Learning

The new cognitive components have been added into the PMFServ architecture and interact with the objects and data utilized by this architecture. In PMFServ,

Figure 5.4: Cognitive Model Connections



agent actions generate events. A typical event in PMFServ includes at least an actor (agent initiating the event) and an action that the actor is performing. It may also include additional information such as a target of the action or results from the action. For a standard PMFServ agent, all observed events are processed fully and the agent's emotional response to them is stored.

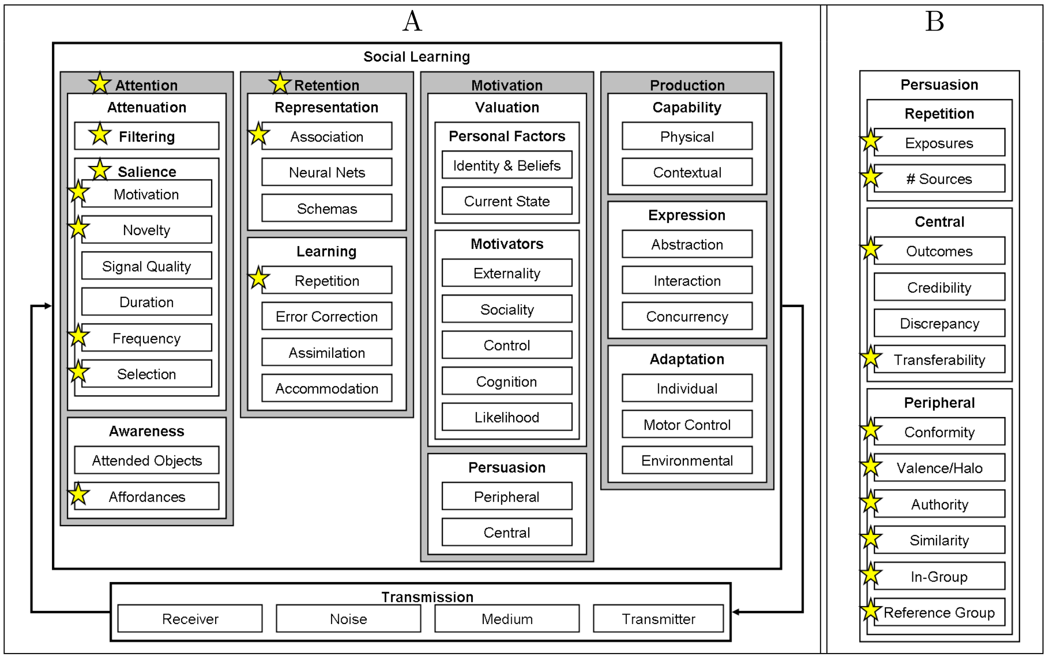
Table 5.2: Standard PMFServ Event

Actor	Action	Target	Result

The new cognitive components allow agents to filter the events that they perceive based upon attention constraints, typically a more realistic scenario for simulating humans. This filtering is done based upon an attention salience factor whose calculation is based on the actor, action, and the agents' emotional response to the event. All events observed simultaneously will be processed by the attention model as a group, which will affect the probability of each event being attended. This is primarily because agents are only able to attend to a limited number of events simultaneously and will tend to miss more events in a busy environment.

If an event is attended, it will reach the learning and memory models.

Figure 5.5: Meme Conceptual Model (Implemented)



These models determine if information about the event is retained and which information is retained. Learning is determined by a fully random learning factor which randomly decides which events are stored. The memory model stores all such events in an associative memory structure which can keep track of the number of exposures to any particular entity, such as an agent or an action. The memory model can also keep track of the number of exposures where certain agents or actions were observed together.

Storing copies of actions is important to this model, since the perceptual system is set up to only look for actions which are in memory. This represents the need to remember that one can perform an action before perceiving it as an affordance in the environment. In the current model, such learning can only occur due to observational learning. In real life people sometimes infer the existence of novel actions, but that functionality was not necessary for the memes of interest in this project.

By extending the agents' cognitive model for processing events and noticing affordances, PMFServ agents have been made meme-capable. If an agent with knowledge of a new action performed that action in front of other such agents, the action could be learned by those agents and imitated. Each new cognitive component added to the PMFServ agent cognitive model extends the agent's ability to analyze events, actions, and agents in their environment.

5.2.2 Social Influence Module

The social influence model set is the first major addition. These models store and calculate social relationships between an agent and its peers. Each submodel calculates a metric for social influence, each of which is used by the motivation mechanism for attention. This allows agents to pay more attention to agents with a high degree of social influence upon them, an important factor in how memes spread within a society.

All factors contained in the social module are relationships toward other agents. The PMFServ standard relationship model implements social factors for valence and agency. Valence is the like or dislike toward another agent. Agency is the level to which the other person is perceived as human and an actor in the environment. The social identity model is a separate model which keeps track of group affiliations, strength of membership, and roles in groups.

Table 5.3: Theories Implemented in Social Influence Module

Theory	Source	Implementation
Dual Process Persuasion	Petty and Cacioppo (1986)	Composite (Partial)
Conformity	Asch (1955)	Direct
Similarity	Platow et al. (2005)	Composite
Halo/Valence	Kelley (1955)	Direct
Authority	Milgram (2004)	Direct
In-Group	Tajfel (1982)	Composite
Reference Group	Kameda et al. (1997)	Composite
Transferability	Bandura (1986)	Direct (Partial)

The social influence theories used to design the social influence model set are noted in Table 5.3[*]. Each of these factors are used as inputs to the attention model, explained in Section 5.2.5. These factors are also used to help mark up an agent's perception of the environment, by adding activations as a function of social influence factors. The mechanisms for applying these additional activations are applied at the scenario level, as part of the perceptual mark ups. This allows them to be processed using the standard decision making algorithm. Each component of social influence will be discussed briefly to explain its contribution.

Dual Process Theories of Persuasion

The dual process persuasion theory has been used as a guideline in the design of event processing but has not been directly instantiated with detailed dynamics of specific dual process models as discussed by Chaiken and Trope (1999).

[*] A credibility component was also designed but was not used during the design process because insufficient data was available to initialize this model.

The dual process theory has been applied to help differentiate between central and peripheral factors, a key aspect of Petty and Cacioppo (1986) Elaboration Likelihood Model (ELM). Within the context of processing an event, central processing evaluates the actions and outcomes involved. This is because an event's action is activity (signal) that an agent sends into the environment. Aspects of the agents involved may also be considered, but will not be considered in isolation. Peripheral factors within the PMFServ cognitive model are those which depend only on the relationship between the perceiver and the observed agent.

The social components are primarily peripheral factors, with respect to attention and persuasion. Conformity, similarity, valence influence, authority, ingroup membership, and group reference value are peripheral cues used by the agent cognitive model. Purely central cues include novelty and repeated exposures. These cues will be discussed in the section on the attention module (5.2.5). Motivated attention, selective attention, and transferability have peripheral and central components, so these are counted as central processing. While these classifications are not stated at the code level, they provide interesting semantic considerations for how these components contribute to the cognitive model.

Conformity Influence Model

The conformity model has its theoretical roots in the seminal work done by Asch (1955). Later work by Tanford and Penrod (1984) proposed the Social Information Model (SIM), a probabilistic conformity influence function. Using this function, a curve was derived for conformity influence based upon upon the number of conforming agents and the number of dissenting agents. This two-input function was used as the basis for the conformity influence model added to PMFServ. The Tanford and Penrod (1984) analysis produced a curve as stated in Equation 5.1, where S is the number of influence sources and T is the total number of targets (naive agents that are not influence sources).

$$ConformityInfluence(S,T) = e^{-4*e^{\frac{-S^{1.75}}{T}}} \tag{5.1}$$

The implemented conformity model uses this equation verbatim. However, the context of its usage is slightly different than that of the original SIM model. While that model assumed a set of confederates, these models assume agents act based upon their own opinions but still exert influence. As such, any set of agents engaged in a particular activity form a group of influence sources (S). The remaining agents involved in other activities are the target group (T).

As such, agents can calculate the conformity influence of any activity in the simulation as a function of the number of agents it sees engaging in the action versus those who are not engaged in the action. This conformity term is then passed to other models such as attentional salience, to determine how an agent learns and behaves.

Similarity Influence Model

The similarity model calculates a social influence factor based upon how much an agent feels it has in common with another agent. The influence of similarity on attention and influence has been an influential topic in the domains of social psychology and social network analysis (Platow et al., 2005). In a real social environment, this type of influence is quite complex due to the subjectivity and iterative nature of determining who is similar to oneself. Perceptions of similarity are based off of behavior, social cues, and secondhand knowledge.

Normatively, similarity between beliefs helps agents determine who is likely to want to engage in similar behavior, such as common interests. Work using PMFServ has approached this issue, by attempting to build models of others' beliefs using parameter estimation approaches (Johns, 2007). However, this work used primarily normative approaches such as simplex optimization and would not be appropriate for this project, which attempts to descriptively model humans. However, modeling how humans estimate similarity in a descriptively detailed way is a complex issue. Social cues such as clothing, speech, and other commonly studied metrics are too fine grained for the scope of this project.

PMFServ contains a second model for estimating similarity, known as GSP congruence (Silverman et al., 2006). This model assumes that agents accurately perceive the similarity of other agent's personalities with respect to their own. As noted, the GSP model in PMFServ is a hierarchal set of weighted nodes. In order to calculate GSP congruence, the two agents' GSP trees are transformed into vectors of normalized linear weights. Each element of these vectors represents specific personality trait. GSP congruence is calculated as a distance between these vectors. The standard GSP congruence function is shown in Equation 5.2, where $\overrightarrow{W_{GSP1}}$ is the perceiving agent's GSP vector, $\overrightarrow{W_{GSP2}}$ is the observed agent's GSP vector, and N is the number of elements in $\overrightarrow{W_{GSP1}}$.

$$GSPCongruence(\overrightarrow{W_{GSP1}}, \overrightarrow{W_{GSP2}}) = \frac{\sum_{i=1}^{N}(\overrightarrow{W_{GSP1}}[i] - \overrightarrow{W_{GSP2}}[i])^2}{\sum_{i=1}^{N}(\overrightarrow{W_{GSP1}}[i])^2 + (\overrightarrow{W_{GSP2}}[i])^2} \tag{5.2}$$

By allowing agents to detect this factor without noise, the model assumes that the agents generally estimate an accurate perception of similarity. This is done by allowing agents to directly access another agent's GSP (Goals, Standards, and Preferences) in order to calculate a similarity metric. This model is typically applied where agents are expected to have a priori knowledge about other agent's personalities, such as agents who have know each other well. Even where agents are not familiar, it provides a useful first order estimate of the perceived similarity which is appropriate when people can quickly generate accurate perceptions of similarity.

The similarity influence model builds off of the GSP congruence model, using GSP congruence as a similarity term. This model also operates as a wrapper to allow subclassing the similarity influence functionality, as not to make it wholly dependent upon the specific implementation of GSP congruence.

Halo/Valence Influence Model

The valence influence model represents the social influence caused by general like or dislike of another person. This is related to the "halo effect," such as where an attractive person appears more competent (Kelley, 1955). Experiments such as Hilmert, Kulik, and Christenfeld (2006) have experimentally shown that valence can affect social influence. Since PMFServ already has a model for maintaining valence, the valence influence model consumes and exposes this parameter so that it can be exposed as an influence value. Since valence ranges from [-1,1] in PMFServ and all influence values are fitted into a range of [0,1], a small transform is applied to valence values to rescale and shift it into the appropriate range.

Authority Influence Model

The authority influence model represents the additional influence conferred by a position of authority. The effects of authority on behavior have been well documented by Milgram (2004) and Mantell (1971). PMFServ has the ability to represent the authority of agents within the groups which they belong to (Silverman et al., 2006). This value fits within the appropriate range and has the appropriate semantic meaning, so the authority influence model wraps and exposes this authority influence metric so that other models can take this into account.

In-Group Influence Model

The in-group influence model represents the social influence based on belonging to a mutual group or clique (Tajfel, 1982). PMFServ has a structure for representing group membership, which allows members to be part of a group. Similarly, groups can be arranged in hierarchies that confer membership into supergroups. The current implementation of in-group influence counts an agent as belonging to the same in-group only if they share the same primary group. This means that while in-group influence is technically a value with a range of [0,1], it actually functions as a boolean value.

Reference Group Influence Model

Reference group influence represents the influence based on an agent belonging to a group against which an agent compares themself, such as a desirable group (Kameda et al., 1997). PMFServ has an analogous factor within its model set that is an agent's "internal membership" with a group (Silverman et al., 2006). Internal membership measures how much an agent desires to participate and support a group. Since this measure is explored within other papers, it will not be covered in detail here.

Reference group influence uses a variant of PMFServ internal membership that has been scaled to fit into a range of [0,1]. This model can report back the desire to belong in any given agent's group (if they belong to a group). This approach to reference group influence has a few important dynamics of note. Firstly, the value can cover anywhere in the range of [0,1], unlike in-group influence. This value is also independent of in-group influence, in general. However, the calculation of internal membership bases some of its parameters upon the perceived leader of the group being considered. This leader is either a specifically designated agent, or the agent with the highest authority. In particular, the GSP congruence of the leader is used as a metric similarity with the group as a whole. This means that reference group influence and similarity influence will have some covariance when an agent considers a leader, since the leader's parameters represent himself and are partially representing his group's influence.

Transferability Influence Model

Transferability influence refers to the additional influence conferred by an agent who has similar capabilities and does actions that one could imitate. Often, this trait is studied in children at different developmental stages. Children have a preference to attend and imitate those of similar ability level on tasks (Bandura, 1986).

The transferability influence model allows agents to process an observed event and determine if they could do the same action at the current time. This determination is only based upon the agent's current affordances at the particular moment, not any past or potential affordances. This implementation has the advantage of easily classifying events into those which they could imitate and those that they could not. However, it is conservative since agents will not assess an action as transferable (imitable) if it is not currently available- even if they did that action previously. With that limitation in mind, this implementation still allows the agent to consider important information about their ability to imitate an activity and consider that when processing events.

5.2.3 Memory Module

The memory module implemented for this project is a simple associative structure that implements encoding, storage, and retrieval processes. Table 5.4 notes the theories used to construct the memory module. Associative memory works by strengthening connections between elements stimuli or constructs due to repeated pairing (Mackintosh, 1983). Association can be considered the strength of connection between two elements in memory. The memory module is also intended to model familiarity, which is generally trained using repeated exposures of the same stimuli or object. In this way, familiarity can be thought of as the strength of a particular element in memory.

Table 5.4: Theories Implemented in Memory Module

Theory	Source	Implementation
Associative Learning	Mackintosh (1983)	Direct (Partial)
Exposure Familiarity Rate	Bornstein (1989)	Direct
Emotion Tagging of Memory	Canli et al. (2000)	Direct (Partial)

This memory model implements both of these processes using a very simple storage mechanism that keeps a record of the events that it has learned. During the encoding process, each event is broken down into its constituent parts (i.e. action, actor, target, result). Every encoded event has a unique set of entities occupying each role. The model keeps a count of how many times each unique permutation has been stored. Using this data model, it is possible to calculate the number of recorded exposures to any individual entity (ex. an agent), the number of recorded exposures to any unstructured set of entities (ex. agent1 seen with agent2), or the number of recorded exposures to any structured set of entities (ex. agent1 hit agent2).

Memory Encoding

The perception process passes currently observed events to the memory model for encoding. Before any events are passed to the memory model to be registered, they must first pass through the attention model (explained in Section 5.2.5). Attended events reach the memory model, which first passes them to the learning model. The learning model determines which events an agent stores and what it stores, which is explained in Section 5.2.4. As a result, the learning model handles much of the encoding process. After passing events through the learning model, the encoding process adds additional metadata that tags the event with an emotional valence and initializes a time value that tracks the age of the event. Emotional tagging was added to assist with recall functionality, and evidence suggests that such tags are an important mechanism in memory (Canli et al., 2000). After encoding the event information, the storage mechanism handles the encoded information.

Memory Storage

At the conceptual level, the memory storage mechanism of the memory model is very simple. Encoded information is stored inside a single entry for each unique set of event information (which will be referred to as a memory pattern). The content of each memory pattern entry is defined by the learning model. The learning model stores the data contained in a standard PMFServ event: actor, action, target, and result. For a valid entry, at least one of these fields must contain data, but not all fields have to be defined. This allows the learning model to store partially processed events where only part of the event was encoded for storage as a memory pattern.

For each memory pattern, the memory model stores a number of exposures, a valence, and an age. The number of exposures for an event increases by one each time an event is stored that matches this memory pattern. When this occurs, a valence is calculated for the event based upon its activations (emotional outcomes). This positive or negative valence is added to any prior valence toward the pattern. Finally, whenever a new exposure is added to the memory pattern, the age of that pattern is reset to zero. As simulation time passes, this age is incremented to keep track of the length of time since a particular pattern was observed. This provides a recency metric for the memory model to use for recall.

Memory Retrieval

The memory model supports two types of retrieval which can be used by an agent: familiarity and unprimed recall. Familiarity represents an agent's sense that an object or action is well known. Within the implemented model, familiarity is calculated as a function of the number of stored exposures to an entity, counted across all memory patterns containing that entity. The familiarity calculation does not consider the role of an agent within a memory pattern, as research does not seem to indicate that such a distinction is important. The familiarity equation is stated in Equation 5.3. The input to the equation, *Entity*, is an action, agent, or other entity contained within a learned pattern. N_E is the number of exposures to that entity and r_f is a familiarity rate that determines the steepness of the curve. Within the current implementation, r_f is set to 0.2 as this allows familiarity to reach 95% after 15 exposures. Empirical research indicates that the exposure effect hits its maximum after between 10 and 20 exposures, so this seemed to be a reasonable familiarity rate (Bornstein, 1989). Familiarity is used by the novelty model, which has important effects on attention.

$$Familiarity(Entity) = 1 - e^{-r_f * N_E} \tag{5.3}$$

Unprimed recall deals with the process of an agent producing a pattern in memory without any particular cue to tell the agent what to think of. This is the opposite of primed recall, which would be where a cue is given and an agent produces a response. Unprimed recall is handled by setting up a function that establishes a recall weight, based upon the valence and age of an event. Equation 5.4 shows the recall weight function, where Val refers to the valence of a pattern and Age refers to the amount of time units (in simulation steps) that has passed since the that event was last stored. Of all patterns, the one with the highest recall weight will be recalled during unprimed recall.

$$RecallWeight(Val, Age) = \frac{|Val|}{1 + Age^{-e}} \tag{5.4}$$

The unprimed recall equation is not directly used by any other models, but can be used during agent actions. It is designed to recall a recent, emotionally charged event. It incorporates some of the insight about emotional tagging on retrieval (Canli et al., 2000). However, the specific values of the recall weight function have no theoretical validity and are simply a heuristic for selecting a memory in an unprimed context.

5.2.4 Learning Model

The learning model is rudimentary and designed only for basic learning of affordances by random discovery or observation. The conceptual models for retention are not vital for this study of affordance discovery. This learning model is sufficient for these scenarios, since the outcomes of the affordance will be constant and transparent to the agent. The expression model has been kept simple also, assuming that the only form of transmission possible is performing the afforded action.

The conceptual models involved in the learning model are noted in Table 5.5. Perceptual learning is not as richly modeled as stated in E. J. Gibson and Pick (2000). This implementation of differentiation learning only recognizes unknown actions from known actions, while accommodation processes only check that an agent can perform an action. These are sufficient this study of memes, however.

Table 5.5: Theories Implemented in Learning Model

Theory	Source	Implementation
Affordances	E. J. Gibson and Pick (2000)	Direct (Partial)
Repetition Effect	Ebbinghaus (1913)	Emergent (Partial)

The learning model receives knowledge of events passed to it by the perception model, which filters events using the attention model. The learning model processes events and determines what is learned from each event. This model takes a simple approach of giving each event an independently random chance of being learned. The model is given a static, global learning probability for each event. If an agent learns an event, they learn the actor and the action from that event. This keeps the learning model simple and allows attention to drive the dynamics for retaining memes. However, the learning model can still handle dynamics such as the Ebbinghaus (1913) learning curve as needed.

The learning process maintains a memory of affordances within the scenario. These affordances provide types of possible actions for an agent. By limiting agents' actions to the affordances that they are familiar with, agents can socially learn that they are afforded certain actions. This learning allows memes to reproduce within the scenario.

5.2.5 Attention Module

Returning to the conceptual model, it is clear that perception is a key gateway within the model. The standard perception model handles awareness- a listing of entity and affordances. An agent's stress level, physiology, emotions, and learning all affect this vital process. By default, a PMFServ agent perceives all the entities and affordances in its environment and evaluates them. The attention model places a filter over this process, limiting the number of entities and actions that can be evaluated. It also calculates an attentional salience factor. This salience function calculates the level of attention focused on some other agent performing an action which has certain results. This salience determines the probability that an event in the environment will be noticed. A noticed action has the ability to generate emotions and to allow an agent to learn a new affordance.

Table 5.6: Theories Implemented in Attention Model

Theory	**Source**	**Implementation**
Affordances	J. J. Gibson (1986)	Composite
Novelty	James (1890)	Direct
Repeated Exposures	Ray and Sawyer (1971)	Emergent
Selection	Simons and Chabris (1999)	Direct
Motivation	Fazio et al. (1994)	Composite (Partial)
Salience	Treisman and Gelade (1980)	Composite (Partial)

The attention model is a composite of smaller models implementing the subcomponents of salience. The constituent attention theories for this model are displayed in Table 5.6. Salience determines the likelihood of observing an event, relative to other events occurring simultaneously. Submodels for attention calculate the motivation, novelty, and selection factors for an event. Each of these models is implemented only to handle attention to semantics- it is assumed that the physical properties (syntax) are equally noticeable. Models for signal quality, duration, and frequency are not implemented at this time.

Novelty Model

Novelty is a theoretical construct that indicates how "new" a stimulus appears (James, 1890). Novelty and familiarity would seem to have an inverse connection, with respect to exposures (Johnston, Hawley, Plewe, Elliott, & DeWitt, 1990). To harness this, the novelty model accesses a record of the number of exposures for each action and agent over time and calculates a novelty factor based on the level of familiarity. The novelty model accomplishes this by reading from the memory model, which has functions to count the number of exposures and to calculate a familiarity value. This familiarity value will be explained later in the section on memory models, Section 5.2.3. For any given event, the novelty is calculated as the RMS of the familiarity values of the actor of the event and the action of the event. The novelty calculation for an event is shown in Equation 5.5, where f_{Actor} is the familiarity of the event's actor and f_{Action} is the familiarity of the event's action according to the memory model.

$$Novelty(Event) = \sqrt{0.5((1 - f_{Actor})^2 + (1 - f_{Action})^2)} \tag{5.5}$$

This representation was chosen because it allows a high degree of novelty if either component is novel. This dynamic was chosen because it allows representation of processes such as dishabituation, where adding an additional stimulus can restore responding to a habituated (familiar) stimulus. In this context, the response of interest is active attention. This implementation allows a return to novelty when a highly familiar person suddenly engages in a totally new action. Conversely, if a straight average was used, then a completely familiar person could be at most 50% novel. Alternatively, taking the maximum novelty component would go too far in the opposite direction: giving no additional novelty to a new person doing a new action as opposed to a new person doing an old action. While a root mean square may not be the best representation for combining these terms, it parsimoniously represents these important dynamics within the simulation.

Repeated Exposures Model

Numerous studies have shown the cumulative impact of multiple exposures and repetition on the cumulative likelihood of attention and impact of persuasive messages (Ray et al., 1971; Ray & Sawyer, 1971). From Ray and Sawyer (1971), it can be observed that across experiments the recall probability of a message tends to have its highest increase with the first exposure. The next 5 subsequent exposures to an advertisement have less impact and tend to either have equal impact (linear curve) or decreasing impact (sigmoidal). The next exposures tend to either result in nearly full recall (hit the upper bound) or have minimal contribution to recall. The Ebbinghaus (1913) learning curve takes on a sigmoid-type function, so this is assumed to be the family of curves that repetition takes on with respect to recall (due to some combination of attention and learning). The persuasive impact of messages is a more complicated issue because it appears to be a function of the persuasiveness of the message. Some messages appear to have little impact, regardless of the number of exposures, while others increase as a function of exposures. This seems to indicate that the impact of repeated exposures is dictated by processing of the content, and not necessarily due to familiarity with the message.

While these represent an increased cumulative impact, empirical studies do not indicate repeated exposure effects that cannot be otherwise explained by other cognitive components. The persuasion of a message appears to be largely dictated by its content and processing, while learning it is modeled by other parts of the agent cognitive model. As such, no explicit repetition model was implemented since its key dynamics are present in the memory model and the novelty model. The memory model captures a record of attended and stored exposures for each agent. The novelty model provides a decreasing impact for each additional exposure, capturing one typical dynamic of repetition on learning. Through these dynamics, the effects of repetition should emerge: greater total familiarity with the presented message and decreased impact of additional exposures.

Selective Attention Model

Selective attention is a construct that refers to the additional probability of perceiving events performed on an object that an agent actively perceives, as opposed to other peripheral events (Simons & Chabris, 1999). Selective attention is implemented by having agents keep a record of the objects and agents they are actively attending at the current time. PMFServ agents are able to actively take actions on other agents, including actions of active perception (watching). As such, the selective attention model records all entities that an agent is currently engaged in action upon. This means that selective attention is focused on any targets being watched or acted upon by an agent. This allows agents to choose who will be the target of their selective attention, as is observed in the cocktail party effect (Cherry, 1953).

$$SelectiveAttention(x) = \begin{cases} \frac{1}{N} & \text{if } x \in X_{Targeted} \\ 0 & \text{if } x \notin X_{Targeted} \end{cases} \tag{5.6}$$

If an agent is allowed to engage in multiple actions simultaneously, their total selective attention is spread evenly across those objects. Equation 5.6 displays the selective attention focusing calculation, where $X_{Targeted}$ represents the set of all entities targeted by an agent's actions, N is the number of entities in $X_{Targeted}$, and x is some entity from the simulation. At present, no mechanism exists to preferentially apply selective attention to certain agents or objects. In the simulated scenarios explored later, agents are only able to engage in one action at a time so selective attention will always be fully focused on one entity.

Motivated Attention Models

Motivated attention is a construct that refers to the additional attention given to events that correspond with the needs, wants, and other motivations of an agent (Fazio et al., 1994). Motivation is the most complex submodel of salience. It calculates a motivation factor based upon the characteristics of the action as compared to the agent's current state. Motivation has two components in this implementation: outcomes (central) and social (peripheral). The outcomes from the action can be motivating, such as seeing someone eat when you are hungry. The social component would be the motivation to watch someone eat because you enjoy their company. Outcome motivation is calculated as a congruence between the agent's current needs on their GSP and the activations from performing the action. The social components use social influence terms which have already been discussed earlier (conformity, similarity, valence, authority, in-group, reference group). All factors of motivation are taken as having an independent impact, following the design decision to keep the model simple where empirical interactions are unknown.

The central motivational cues are handled by allowing agents to analyze the outcomes of events which have occurred. As noted earlier in Section 5.2, agents evaluate their potential actions based upon "activations" that determine the attractiveness of that action, as mediated by their values and beliefs. To calculate a factor for motivated attention, an agent processes an event that results from some other agent's action. In processing this event, the agent calculates the subjective emotional utility for themselves had they been the actor in that event and the outcomes were the same. So, for example- if agent B is eating a sandwich, the motivational salience for agent A is a function of the subjective benefit (or harm) for agent A eating a sandwich. This motivated attention does not consider if the action or outcomes of the observed action are possible.

Equation 5.7 displays the central motivated attention calculation for an agent observing a given event (Note: the 'sgn' symbol represents the sign function, producing -1 for negative values and 1 otherwise). SEU_{Event} represents the subjective expected utility of activations that the perceiving agent would receive had they been the actor in that event and the outcomes were the same. Two adjustments are made to the raw utility value in order to calculate the motivated attention factor. One adjustment rescales the value from between [-1,1] to fit into [0,1].

$$MotivatedAttention(Event) = 0.5 * (1 + sgn(SEU_{Event})(|SEU_{Event}|^{0.25})) \quad (5.7)$$

The second rescaling factor takes the fourth root of the absolute SEU value. This factor was introduced during model calibration due to the very small range over which SEU can realistically operate within PMFServ. An SEU of 1.0 would indicate that an agent went from a completely neutral state to a state of full satisfaction of all its goals, standards, and preferences. In practice, such a huge swing would almost never be observed. This calibration tweak was introduced to spread the range of motivated attention so that smaller changes in SEU would still have some impact on the motivation pay attention to an event. Rescaling was necessary since in experimental studies, even modest changes in motivation such as hunger resulted in significant changes in attention (Fazio et al., 1994). A linear weight was not acceptable, since this would lead to clipping the range of SEU for the purposes of motivation (high motivation and very high motivation would have the same impact). As such, a calibration exponent was calculated from the Stanford Prison scenario which allowed the maximum possible utility changes to span a range between [0.15, 0.85] for the central motivated attention factor. Unfortunately, since motivation does not have a standardized unit or scale, there was no way to calibrate this parameter in a more methodological manner. For a follow up model, this would be an area that would benefit from additional empirical data.

Attentional Salience

Salience is used to calculate the probability that an action is receives enough attention to be processed cognitively. This is accomplished by first calculating a salience for each event occurring during a time step. An additional salience term exists which represents inattention salience: the salience of background events not simulated that might be attended to instead of the simulated events. This vector of saliences is normalized to form a probability vector, from which a finite number of events are chosen. Each event is chosen without replacement, except for inattention which always remains an option. The probability distribution for choosing an event to attend is shown in Equation 5.8, where E is the set of all simultaneously observable events, E_{Att} is the set of already attended events, s_e is the salience of an individual event e, and s_I is the inattention salience.

$$P[e = Attended] = \begin{cases} \frac{s_e}{s_I + \sum_{e \in E \setminus E_{Att}} s_e} & \text{if } e \in (E \setminus E_{Att}) \\ \frac{s_I}{s_I + \sum_{e \in E} s_e} & \text{No Event Attended} \\ 0 & \text{if } e \in E_{Att} \end{cases} \quad (5.8)$$

The algorithm for drawing the set of attended events is displayed in Algorithm 5.6, where N is the maximum simultaneous events attended, E is the set of all simultaneously observable events, and $X(E, E_{Att})$ is a random variable with a distribution defined by Equation 5.8. The output of this algorithm is E_{Att}, the total set of attended events. If an inattention term is selected, it is ignored and one less total event will be attended. This attention algorithm is effectively an iterated drawing from the yet-unattended events, with a constant probability of no event being attended. This corresponds loosely to a series of winner-take-all competitions for attention between events, a process which has some support in neurological research (Lee et al., 1999). These events are processed by the learning model, which can learn new affordances.

Figure 5.6: Attention Algorithm

```
$E_{Att} = \{ \}$
for $i = 0$ to $N$ do
  ATTENDED_EVENT = X(E, $E_{Att}$)
  if ATTENDED_EVENT != No Event Attended then
    $E_{Att} = E_{Att} \cup \{$ ATTENDED_EVENT$\}$
  end if
end for
```

While the parameters used to calculate attentional salience and their basic curves are known, no data exists to define their relative strengths or appropriate combination. To accommodate this uncertainty, multiple classes of functions with different weight parameters are available within the model. By examining the studies that define these parameters as impacting recall of events and/or messages, a linear weight was estimated for each component which represents the slope of change between the high condition and the low condition in the experiment. For example if the high authority condition resulted in a 0.3 increase in probability of recall, this was chosen as the linear weight. Alternatively, for those factors which do have experimentally derived curves (conformity), the curve slope was used instead. All factors were normalized to fit the range [0,1].

Attentional salience is calculated as a function of attention and social influence terms previously defined. These factors are novelty, centrally motivated attention, selective attention, transferability, authority influence, conformity influence, similarity influence, valence influence, ingroup influence, and reference group influence. Each parameter is combined using a linear weight that determines its contribution to the total salience for an event. As such, the attentional salience for an event e is determined by a function as shown in Equation 5.9. The w factors represent the weight given to each factor. This form of equation was chosen as it was the simplest possible combination that would capture the information operationalized from the social science findings and theories.

$$\begin{aligned} s_e = \mathit{Salience}(e) = & w_0 \cdot \text{Novelty(e)} + w_1 \cdot \text{MotivatedAttention(e)} + \\ & w_2 \cdot \text{SelectiveAttention(e)} + w_3 \cdot \text{Transferability(e)} + \\ & w_4 \cdot \text{Authority(e)} + w_5 \cdot \text{Conformity(e)} + w_6 \cdot \text{Similarity(e)} + \\ & w_7 \cdot \text{Valence(e)} + w_8 \cdot \text{InGroup(e)} + w_9 \cdot \text{ReferenceGroup(e)} \end{aligned} \tag{5.9}$$

Table 5.7 notes the weights for each factor, as well as the source used to help initialize these weights. The "Process" column in Table 5.7 refers to if the component is Central (depends on the specific event), Peripheral (depends on more general context), or Mixed (combination of both).

Table 5.7: Event Salience Component Weights

Component	**Assumed Weight**	**Source**	**Process**
Authority	0.33	Mantell (1971)	Peripheral
Conformity	0.34	Tanford and Penrod (1984)	Peripheral
In-Group	0.30	Tajfel (1982)	Peripheral
Motivation (central)	0.47	Roskos-Ewoldsen and Fazio (1992)	Central
Novelty	0.21	Johnston et al. (1990)	Mixed
Reference Group	0.30	Kameda et al. (1997)	Peripheral
Selective Attention	0.32	Simons and Chabris (1999)	Mixed
Similarity	0.47	Platow et al. (2005)	Peripheral
Transferability	0.10	Bandura (1986)	Central
Valence/Halo	0.38	Hilmert et al. (2006)	Peripheral

Each of these weights was inferred from examining the related paper, as noted in Table 5.7. The weights are intended as a "best guess" estimate of the importance of each factor with respect to social learning, due to their observed effect on either attention, perception, or retention. First, the input and output variables of interest were determined. Second, the form of the empirical relationship was determined, to the level of the paper's presentation (ex. correlation, slope, function, etc). The third step was to estimate amount that the input could affect the output, if known. Last, each relationship was normalized so that the input variable ranged between 0 and 1. From these, the salience weights were defined. More information on how these weights were initialized is given in Appendix F. These weights are not intended to be taken as reliable estimates of the relative importance of factors, but were estimated to try to capture major differences between importance of factors.

The limitations to this approach are significant but unavoidable. Firstly, the experiments which prove these factors are important do not generally establish minimum or maximum values for their inputs. Even at the theoretical basis, it is difficult to establish criteria for what constitutes the maximal or minimal level of authority that a person is perceived to have. Secondly, there is no assurance that these factors work linearly or independently. While this attempt at a linear approximation was workable for this research, a better functional combination could be necessary for more in-depth study.

Despite the limitations and caveats to the attentional salience calculation approach, it incorporates the directionality and known functional characteristics of the underlying empirical studies. This provides some insight into how various factors may interact and produces some interesting results that will be noted in Section 7.

Additionally, social learning of affordances is straightforward using this cognitive framework. It requires three conditions: an affordance available to all agents, a set of agents aware of the affordance, and a set of agents unaware of the affordance. When agents choose to perform an action, the OODA loop for each observer evaluates if social learning of the affordance is appropriate. Any affordance in PMFServ can be treated as a meme using this system, without any changes to the affordance.

5.3 Scenario Architecture

Scenario design in PMFServ involves designing the affordances for entities in a scenario. These affordances are rules that determine if an agent can take an action on some part of the environment. These affordances associate with action implementations, which have outcomes that affect the environment and acting agent. Designing the affordances and models for actions creates a family of models which must be populated with data.

Objects are the simplest entities. An object requires data purely to support its affordance functions and the actions performed on it, plus a name and unique id. The initial values for objects are generally part of the assumptions of the model and will not be varied during experiments.

Groups are more complex. In addition to the requirements of an object, groups have membership, resource, and social data. Membership data stores the members involved, their levels of authority, and their roles in the group. Groups may also have subgroups, forming a hierarchy structure. Resource data stores the types and levels of holdings by the group, such as shared economic or security holdings. Economic models in PMFServ make use of these resources, described in (Silverman et al., 2010). These models will only be used for the Iraqi village scenario. Relationship data for groups is limited to a valence toward other groups. The relationships across groups are not always the same as those between agents across groups and should be considered the “official stance” of groups toward each other.

Agents require the largest amount of data. They must be initialized with the strength of social influence factors, weights on the GSP personality tree, learning values, and physiological levels. Social influence data tracks the perceived valence and agency for each agent toward every other agent. The remainder of social data is retained by the group structure and data. The GSP tree requires the largest amount of data, populating the relative weights of a personality tree consisting of dozens of nodes. A knowledge engineering methodology exists for calculating this data through a combination of demographic data and other sources, described in Bharathy (2006) and Silverman and Bharathy (2005). Data in the memory model stores the affordances that an agent initially knows in a scenario. These values are vital to experiments in this model.

5.4 Simulation

A PMFServ simulation occurs in discrete time and can be considered as a Markov Decision Process (MDP). PMFServ simulations support discrete state models or continuous state models (to the level that computer simulation allows). The transitions between states are determined by the actions performed by agents. Agent action is simultaneous- all agents make their decisions based upon the present state. The set of an agent's decisions is their chosen behavioral expression for that time step. The transitions for the system are a function of the full vector of agent decisions, which may have covariant effects.

State transitions may be random or non-random, depending on the nature of the actions available and environmental effects added by the simulation. Simulation effects are generally minimized but may be required to resolve resource conflicts, such as two agents intending to use the same door. Randomness in action implementation and covariance with other agent actions cause the difference between decisions and behavioral outcome in the environment.

The two scenarios implemented in the next section have different simulation setups. In the Stanford Prison Experiment scenario, all agents act simultaneously. This means that each agent generates an event that competes for attention on each step. In the Iraqi village, all agents are allowed one action per time step, taken based on a turn order. The turn order is unimportant, due to the scenario designs. In this case, the events created by agents only compete against inattention salience (as defined in Section 5.2.5). To account for this, the Iraqi village has a higher inattention salience value which represents the greater level of ongoing activity that is not simulated. In both scenarios, no simulation effects are applied other than performing the agent actions and applying the effects. Randomness is introduced by the attention model, which probabilistically attends to events.

The Iraqi village differs slightly in its implementation. This scenario has an additional source of randomness, where the meme expression has different probabilistic effects. It also employs actions that occur over time. This is accomplished by disallowing an agent from choosing a new action until the ongoing action is terminated or suspended. While slightly different in sequencing, the same models apply for this simulation.

Chapter 6

Experiment Design: Affordance Discovery

Two test scenarios were selected for experimentation, each with calibration and validation goals. These scenarios employ different types of data and different standards of evaluation specific to their purpose. The first scenario was an analog of Zimbardo's Stanford Prison experiment, examining the spread of suppressive and rebellious actions (Haney, Banks, & Zimbardo, 1973a). This experiment included a training stage used for hand tuning of model connections. The Stanford Prison experiment was also used for a selection of internal and external validity tests. The external validity test for the Stanford Prison simulation is designed to validate the transmission dynamics of the model. The second scenario was intended to analyze competition of memes in a complex environment. This scenario was built using human terrain data for an Iraqi village provided by the United States Marine Corps.

Experiments were performed in four steps: scenario design, initialization, simulation, and analysis. The scenario design phase involved designing the actions and entities present within the experiment. The initialization phase involved estimating initial state values for the data of the model. Before simulating, each experiment was assigned a set of experimental cases. Setting up experimental cases involved selecting the independent variables that would vary between different experimental cases. Each independent variable for an experimental case would be assigned an initial value for that case based upon either a static value or a random variable. As such, the simulations differed only by the distribution that generated their initial state prior to simulation. Since memes are the focus of interest, the agents initially aware of each meme were used as the independent variable.

During the simulation stage, each experimental case was used to generate a set of runs. A run is a specific number of simulation steps that defines a state trajectory (the path of states the simulation ran through). The number of steps were calibrated for each scenario to ensure that a majority of agents have the opportunity to become aware of the meme before the run halts. Since PMFServ runs as a discrete-time simulation, each time step must be assigned units that determine the amount of time that passes between ticks. The time interval assigned to a time step has an impact on time-sensitive functions such as physiology (ex. hunger), activity context (job shifts), and the decay rate for emotions. The level of granularity required for modeling affects the length assigned to time steps. Additionally, the number of steps for a run was also bounded by simulation runtime and data storage concerns. Precision in modeling was balanced against hardware concerns, both for simulation and analysis.

Analysis of data was conducted using established statistical analysis tools where possible, as well as developing a novel analytical tool for comparing ordered sequences. The internal validity analyses were expected to reproduce results consistent with the underlying empirical research used to generate the cognitive models in Section 5.2. These measures also produced some unexpected results that have interesting connections with the social science theories used to generate the computational model. External validity testing was done by comparing experimental results against hold-out data that was not used for experimental calibration. External validity metrics were only available for the Stanford Prison Experiment scenario, since the Iraqi Village scenario did not have behavioral metrics to use as a comparison. In additional to examining each model in isolation, cross-validation between the experiments was conducted where possible.

6.1 Scenario 1: Stanford Prison Experiment Simulation

The first scenario was a classical social science experiment, chosen to help calibrate the connections between models and perform external validity checks. The Stanford Prison Experiment case study (Haney et al., 1973a) was translated into a scenario, to help examine the effects of social connections, groups, and roles on meme transfer. This scenario design proved workability, that the model can be implemented and studied.

The Stanford Prison Experiment was conducted in 1971 and was intended to explore of the impact assigned roles had on behavior inside a simulated prison environment (Haney et al., 1973a). In the experiment, 24 subjects were selected out of a group of 75 applicants based upon their psychological test results which indicated they were mentally stable and that their scores were relatively close to "normal" (i.e., the mean of the tests). These subjects were randomly assigned to be prisoners or guards. The experiment, intended to last two weeks, lasted only 6 days due to the growing abusiveness of the guards and signs of distress among the prisoners. Haney et al. (1973a) interprets this outcome as evidence for the role of situational factors in causing institutional abuse, as opposed to purely individual factors.

The conclusions of the study have been contested since its publication, with a variety of alternate hypotheses suggested for the causes of cruelty within the prison. Carnahan and McFarland (2007) presents data which suggests that self-selection may have given a disproportionately cruel subject pool, since the call for subjects noted it involved prisoner and guard roles. Fromm (1973) and others have suggested that the since the guards were not uniformly cruel, individual factors were still a major driving force for abuses. It has also been suggested that a major cause for the abuse of prisoners was the orientation given to guards, in which they were informed that part of the intent of the prison was to make the prisoners feel powerless (Reicher & Haslam, 2006). The intention of using this scenario is not to assert a position with respect to the cause of all abuses within the experiment, but to explore the possibility that social learning played a role in how certain abuses and resistance unfolded within the experiment.

The Stanford Prison Experiment was chosen as a scenario to model because it was a controlled field study which collected a data using a variety of collection methods. The Stanford Prison Experiment researchers collected data that included personality traits, emotion surveys, social groups, and detailed behavioral logs. Despite the Stanford Prison Experiment's status as a controversial study, there simply have been few studies released that have this breadth of data.

6.1.1 Stanford Prison Experiment Data Sources

Data for the Stanford Prison Experiment was collected on site at the Archives for the History of American Psychology (AHAP), under special permission from Dr. Zimbardo and the AHAP archival staff. All data from the experiment was present only in print, with some holdings of the archive present only in raw form (no reliable subject code keys). A week was spent working with the archive staff to collect redacted and subject-coded papers and data from the archives, according to a code key developed for this project. Certain data from the archives was missing or only partially complete, but the total quantity of information in the holdings related to the Stanford Prison Experiment was large and very useful for setting up a meaningful scenario for simulation.

The data extracted from the archives included qualitative and quantitative information. Table 6.1 displays the types of data available from the Stanford Prison experiment. As is common in dealing with archival data, each of these data sources had some missing data. In some cases the missing data was incidental, while in some cases the raw data no longer existed and only metrics on the data had been archived.

Table 6.1: Stanford Prison Experiment Information

Data Source	**Use For Simulation**
Comrey Personality Inventory	8 factor personality trait inventory
F-Scale	Authoritarian personality measure
Mach Test	Measure of machiavellianism
Mood Adjective Checklist	Measure of positive and negative affect
Action Frequency Metrics	Frequencies of actions occurring (coded from video)
Hour By Hour Logs	List of recorded events, with approximate times

Personality trait information is available through the Comrey Personality Inventory (Comrey, 2008), the F-Scale (Adomo, Frenkel-Brunswik, Levinson, & Sanford, 1950), and the Mach test (Christie & Geis, 1970). Personality data from each of these inventories is contained in Appendix H. The Comrey inventory is the most comprehensive measure of the measures, consisting of 180 questions which are used to derive metrics for 8 traits: Trust, Orderliness, Conformity, Activity, Stability, Extroversion, Masculinity, and Empathy. While the Comrey Inventory has been used less frequently since the Stanford Prison Experiment, studies continue to examine the constructs involved- especially in comparison to other trait inventories such as the Big Five factors and the MMPI (Paunonen & Jackson, 2000; Rushton & Irwing, 2009). The raw data was available for only 3 subjects, all of them prisoners. However, an intermediate form of data existed which listed each subject's standard deviations from the mean trait value of their group (guards or prisoners). Finally, the mean and standard deviation for each trait was available for each group. Given the level of precision involved in the model, the mean and standard deviation data was sufficient to estimate the personality trait differences between agents. Additional information about this process is found in Appendix H.

The F-Scale measure, though intended to measure authoritarian tendencies, has been shown to be a better indicator of racist tendencies and a tendency toward in-group centric attitudes (Eckhardt, 1988). A similar set of partially complete data was available for the F-Test results. 9 prisoners and one guard had raw F-Test results available. As with the Comrey results, the mean and standard deviation of the results for each group was available for the F-Test. The F-Test data for guards would have been a loss, except for the fact that the guards had a relatively low variance on this measure, with $\overline{x}$=4.36 and s=1.19. Since the one known value was an 8 it accounted all of the variance, this meant that all other guards scored exactly 4.

The Mach tests recorded a measure of the Machiavellianism of subjects, in terms of their willingness to use others as a means to an end (Christie & Geis, 1970). For the Mach tests, the mean and variance of each group was known but the specific values were incomplete. However, as with the prior measures, the subjects whose data was recorded were the deviants- subjects who varied significantly over or under their population mean. This mean that the agents without specific data could be constrained into a fairly narrow range. The same approach applied to missing measures on the Comrey inventory was applied to the Mach test, and is documented in H.

The Mood Adjective Checklist (MAC) questionnaire was filled out by prisoners and guards at three separate time points throughout the experiment, each about 2 days apart. The MAC questionnaire measures positivity, negativity, activity, and passivity for a subject for the moment the survey is filled out. It was expected that the raw data for the MAC measures would provide additional information about the emotional state of specific agents over time, or at least provide the specific data points for emotional trends for each group (guards and prisoners). Unfortunately, no MAC data remains that explains the specific emotional state of individual subjects. Worse, some of the original questionnaires were either lost or never filled out. This means that while means and variances are present for each questionnaire, these aggregate values are missing data. Different sets of subjects are missing from each questionnaire, meaning that mean values for prisoners are not comparable even between the first and second batches of questionnaires. Though the raw data was not available, published papers from the experiment reported emotional trends of the prisoner and guard groups (Haney et al., 1973a). Given that even the original researchers were missing data however, the emotional trends reported in papers such as Haney et al. (1973a) must be considered as partially incomplete.

Action frequency metrics were recorded during the experiment by analyzing video recordings taken by inconspicuous cameras during the experiment (Haney et al., 1973a). Approximately six hours of day-to-day activity was recorded during the experiment. Each tape recording was manually coded by researchers, counting the frequency of certain actions over 100 frames of tape (about 6.5 minutes). In addition to counting the total frequency of certain actions over each 100 frames, the total count was broken down into actions directed between different groups (prisoners to guards, prisoners to prisoners, etc). The recorded actions were commands, information, insults, questions, resistance, physical aggression, helping, threats, use of instruments (threatening with a baton), and addressing others individually (individuating reference) or impersonally (deindividuating reference). For each of these actions, some of total data was missing. However, most actions had either a raw count from the daily life tapes or a frequency count. Additionally, almost all actions provided the percentage of such actions that were performed from guards to prisoners and the percentage that were performed from prisoners to guards. A significant amount of action frequency data still existed from the tapes. Since much of the original tapes have been transferred to DVDs in the archives, it might also be theoretically possible to recode this information in order to recapture any lost information. With that said, recoding the tapes was not done because it did not appear to provide much additional data on actions over time and because the poor sound quality of these recordings might make it hard to reliably code speech acts.

The largest data source for the experiment, at least in physical dimensions, is a resource known as the "Day By Day, Hour By Hour Logs." These logs appear to have been compiled by Dr. Zimbardo some time after the experiment as a way to aggregate the events of the experiment into a single resource. These logs are a set of approximately two dozen poster-sized sheets of graph paper, representing the approximate chronology of the Stanford Prison Experiment. Events from each day are presented in approximate order, with either an exact or approximate time listed for the event in a separate column. While not an exhaustive list of the experiment's activities, it captures the key events of the experiment. The log also captures incidental occurrences such as when prisoners resisted and when prisoners were thrown in the hole (a storage closed used for isolation). While on site at AHAP, these logs were manually transcribed into an electronic format- applying a subject code key to remove any identifying information. As a data source, the hour by hour logs were extremely important because they display the order that events occur- an important aspect for studying memes.

Other information about the experiment was collected by examining a transcript of the instructions given to guards and notes about the scheduled activities inside the prison. In addition to these data sources within the AHAP holdings, the published results from Haney et al. (1973a), Haney, Banks, and Zimbardo (1973b), and Zimbardo (2007) based on the experiment were examined closely. These sources gave additional information about the context of the prison environment, which helped in modeling the Stanford Prison Experiment scenario in PMFServ.

6.1.2 Stanford Prison Experiment Scenario Design

Scenario design in PMFServ requires setting up the environment, available actions, and agents that will be simulated. The environment in PMFServ consists of its entities: all objects, groups, and agents within the scenario. The interactions between these entities are determined by the agents' cognitive models, the actions available in the scenario, and simulation settings that determine when agents can initiate actions. The design of these elements will be described briefly.

Entities

As stated in Chapter 5, PMFServ supports three types of entities: objects, agents, and groups. Objects are inanimate entities that do not take actions but may be targeted by actions. Agents within a scenario take actions, typically using an OODA-loop cognitive model to drive behavior. Groups are social structures for agents in which an agent may have membership, roles, and authority. Groups also store collective properties, such as group wealth. The Stanford Prison Experiment simulation uses only agents and groups and does not represent specific objects or locations involved in interactions (ex. doors, food, etc). This approach was chosen because the social dynamics appeared to be the key element of the experiment, rather than the logistics of taking actions.

The Stanford Prison Experiment simulation utilized three types of agents: meme-capable cognitive agents, a minimal PMFServ cognitive agent, and an automaton agent. Table lists of the set of agents present in the simulation, their group, and their relevance to the experiment. All prisoners and guards were instantiated as meme-capable cognitive agents, as described in Section 5.2. These agents were capable of socially-driven attention and could consider social influence on their decision making. All subjects within the simulation are referred to by subject codes established during data collection, since not all participants in the experiment were assigned a consistent code key in the raw materials. Table 6.2 lists the set of agents used for simulating the Stanford Prison Experiment.

The prisoner agents (S_00 - S_09) and guard agents (S_11 - S_21) were represented by meme-capable agents. Subjects S_07 and S_14 were not simulated since they were alternates that did not participate in the experiment. S_00 and S_21 were unique cases because they were alternates who joined the experiment later than other agents, so these late entrances had to be simulated. Subjects S_22 and S_23 were not simulated, since the first was a guard that was present for only one shift, the second was a researcher informant who was present for less than a day. Given that these agents were added later in the experiment, the memes of interest were already prevalent before they arrived. This made their role was minimal for the experiment in general. S_10 was held out of analysis since it unclear if the personality data for this subject was complete.

Prisoners in the experiment were always present once they joined the experiment, for up to 6 days. Guards entered and exited the experiment based upon their shifts. The shifts ran from 10 AM - 6 PM (Day Shift), 6 PM - 2 AM (Evening Shift), and 2 AM - 10 AM (Night Shift). Overlap existed between the shifts where guards tended to interact, partly due to guards staying late and also due to requests from the experimenters to have additional guards present for certain activities.

Since each participant was built from the same template, the basic setup of each agent was relatively similar. Each agent utilized the same model of cognition.

Table 6.2: Stanford Prison Experiment Agents

Agent Name	Group	Importance
S_00	Prisoner	Alternate prisoner added on Day 4 at approximately 7 PM. Quickly resisted and was treated badly by guards in response (force fed, thrown in hole). (Prisoner 416)
S_01	Prisoner	Insulting and sarcastic toward guards, but some resistance. (Prisoner 5704)
S_02	Prisoner	Strategically acted as a "model prisoner" by working hard to obey guard orders. Got the nickname Sarge for his soldier-like attitude. (Prisoner 2093)
S_03	Prisoner	Minor resistance, following others. Released on Day 4 due to stress-related eczema. (Prisoner 3471)
S_04	Prisoner	Resisted early in the experiment, but stopped resisting later. (Prisoner 7258)
S_05	Prisoner	First prisoner to resist, became agitated and was released by the end of day 2 following a revolt. (Prisoner 8612)
S_06	Prisoner	Generally cooperated and did not resist. Felt like he was really imprisoned. (Prisoner 1037)
S_08	Prisoner	Initially cooperated and did not resist until later than other prisoners. Generally kept in mind that the prison was an experiment. (Prisoner 5486)
S_09	Prisoner	In the initial wave of resistance, targeted by punishment. Broke down and was released on day 4. (Prisoner 819)
S_10	Prisoner	Cooperative with no resistance at the start, but resisted occasionally as experiment continued. (Prisoner 4325)
S_11	Guard	Night shift guard. Some sadistic and vengeful behaviors, but generally just played guard role.
S_12	Guard	Night shift guard. Attempts to be "stern, but not overzealous" but regularly degrades prisoners. Appears to take a leadership role in his shift.
S_13	Guard	Evening shift guard. Referred to as "John Wayne" he is the most verbally abusive guard and is noted as an innovator of sadistic punishments. De-facto shift leader.
S_15	Guard	Evening shift guard. "Good" guard who attempted to treat prisoners fairly, avoiding severe or arbitrary punishments. Initial power struggle with S_13.
S_16	Guard	Day shift guard. Did not tend to abuse prisoners, but did go along with the other guards when needed.
S_17	Guard	Day shift guard. Became de-facto leader of the day shift. Authoritarian and utilitarian approach (enforce rules).
S_18	Guard	Night shift guard. Took on guard role like a job, which became routine.
S_19	Guard	Day shift guard. Avoided harassing or commanding prisoners in general, but did approve of some punishments.
S_20	Guard	Evening shift guard. Harsh and physically intimidating toward prisoners. Followed S_13's lead.
S_21	Guard	Evening shift guard. Added to experiment on 2nd day shift, staying on as a night shift guard.
Experimenter	Experimenter	An agent representing the experimenters. Generic representation of experimenter with authority.
PrisonSchedule	None	An automaton that keeps track of the time of day and controls the scheduled activities for each day.

All agents started the simulation with neutral feelings toward each other, equal physiological states (low hunger, low fatigue), and equal authority within their respective group. This meant that agents initially differed entirely as a result of their personalities, their group assignment, the time they entered the experiment, and their shift (for guards). For this reason, it would actually be trivial to explore the counter-factual case where the role assignments were reversed, making prisoners be guards and vice versa. While this was not attempted, it does highlight one of the advantages of an agent-based modeling approach- the ability to easily explore "what-if" scenarios.

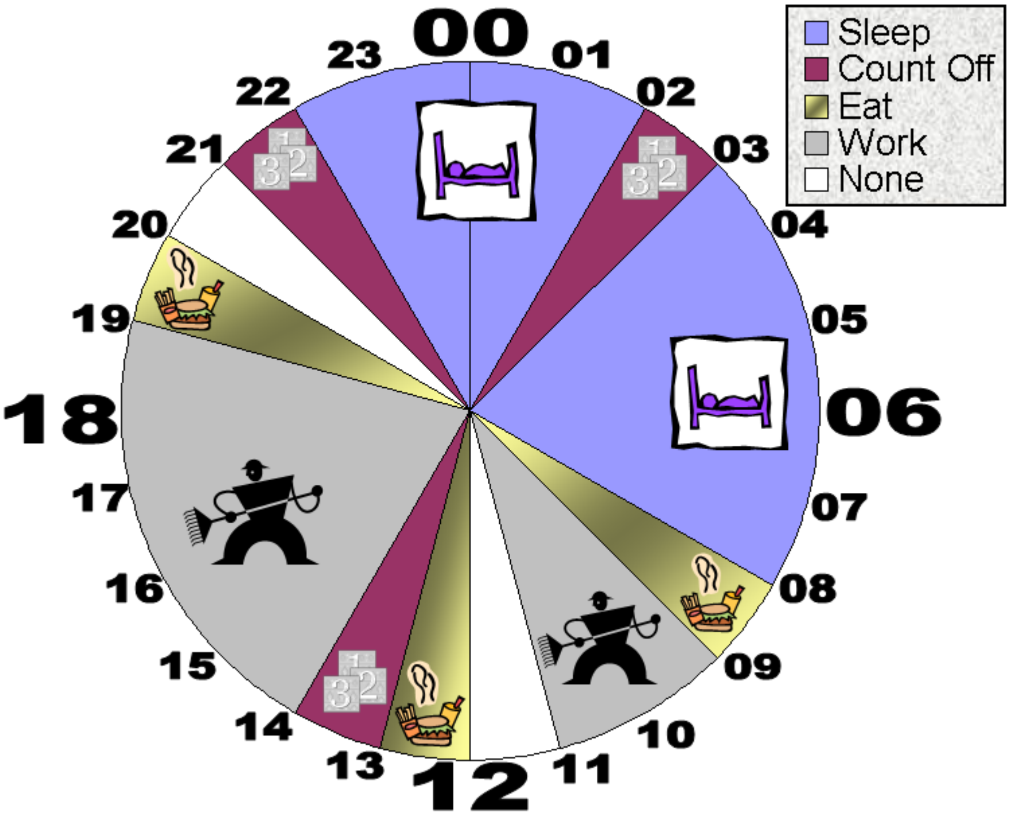
Assigning the guard shifts, group assignments, and time that agents started the experiment was a relatively straightforward modeling task. The richest differences between agents were the differences in the agents' personalities, as represented by the weighting of the GSP trees. Significant time and modeling effort was made to estimate GSP tree weights that captured the differences between agents' personalities. These differences would be a driving force for agent behavior and interaction.

One additional agent exists within the scenario to represent an experimenter observing the experiment. This agent was necessary for two reasons. Firstly, the Experimenter agent was able to dismiss prisoners who had very high levels of stress and were demanding to be released. Since such actions occurred within the experiment, this decision rule allowed participants to be dismissed. Secondly, the Experimenter was used to present memes to participants prior to the start of the simulation. This was intended to represent the Experimenters briefing participants about their ability to perform certain actions a priori and was used for generating one of the experimental cases. The Experimenter agent is not a full fledged cognitive agent in this scenario, however. It has no physiology, agents cannot take actions upon it, and its GSP is fully normalized such that all traits are equally valued. As such, it does not act as a simulated person within the scenario but is present to provide a placeholder for these two pieces of functionality.

Three groups also exist for structural reasons, the Prisoners, Guards, and Experimenters. These groups cannot be targeted by actions but allow agents to have membership and authority within their respective groups. Agents are not given any specific roles within their groups in this scenario, since there are no official leaders or specialists.

Finally, an automaton agent named PrisonSchedule maintains the schedule of activities for the prison and other time-based events. In the Stanford Prison Experiment, guards were expected to run prisoners through a regular schedule of activities. The regularly scheduled activity blocks were sleeping, eating, working, counting off, and unstructured time. Figure 6.1 shows a 24 hour clock which represents an approximate schedule for the typical prison day, with activities rounded to the nearest hour. The PrisonSchedule contained this information

Figure 6.1: Prison Schedule Day



and updated its information so that agents could be aware of the appropriate scheduled activity. This schedule is a slight simplification, since Count Offs occurred with greater frequency but shorter duration during the day, rather than always lasting an hour. It also does not include special activities such as visitors or meetings with researchers, as these periods were not likely times to spread the memes of interest. While the PrisonSchedule agent provides the information to agents about the appropriate activity, it has no direct impact on their actions. For example, if an agent is not hungry at meal time- they may not eat. Except for meal times the schedule does not change the effects of agent actions, it only changes their perception of the situation. However, since both prisoners and guards are aware of the appropriate actions, if a prisoner is not performing the correct activity then guards will have incentive to punish them.

Actions

The actions in the Stanford Prison scenario can be broken down into three groups: baseline activities, transitions, and interpersonal actions. Interpersonal actions were the most important to model because the Stanford Prison Experiment recorded action frequencies for these types of actions. Baseline activities and transitions were modeled to ensure that the right context was present for meme transmission.

Table 6.3: Stanford Prison Experiment Baseline Actions

Action	**Available To**	**Description**
Count Off	All	Agent states their name and number. Modeled as a boring but active speech act.
Eat	All	Agent eats food, if available in the given context. Off duty guards are assumed to have access to food.
Perceive	All	Agent looks at another entity or themselves. Agent only watches and otherwise is inactive.
Sleep	All	Agent tries to sleep. Once an agent is asleep, they remain asleep until they wake or an external event wakes them.
Work	All	Agent engages in repetitive work, such as making beds, moving boxes, or pulling nettles out of blankets.

Baseline activities are actions that an agent can take, even without other agents being present. Baseline activities are available to all agents, even if they are not currently present in the experiment. This allows off-duty guards to handle their basic needs when not in the experiment. This modeling choice was made because while it is impossible to model a guard's life outside of the experiment, it is reasonable to assume that they would eat, sleep, and occupy themselves. As a result, allowing off duty guards to perform baseline actions was more realistic than assuming they returned to the experiment as if the had been in stasis while away. The set of baseline actions is presented in Table 6.3. The majority of baseline activities correspond with the scheduled activities in the Stanford Prison Experiment: eating, sleeping, counting off, and working. While these activities were not directly recorded within the experiment, they were a backdrop interpersonal actions. For prisoners, the schedule determines which activities they should be engaged in. If prisoners are performing these activities during their assigned periods, guards have less incentive to harass them. Conversely, if prisoners hate a particular activity they will be more likely to perform other actions such as resisting the guards.

Transition actions are the simple actions, but are important for the experiment to run realistically. Table 6.4 describes the three types of transition actions: attempting escape, ending shifts, and starting shifts. Attempting escape is available so that prisoners are able to attempt jailbreaks, as happened during the experiment. However, prisoners within the experiment who attempted escape did not generally intend to run away and did this action to gain leverage for negotiating better conditions. For this reason, prisoners attempting escape never actually exit the experiment. Starting and ending a shift are mutually exclusive actions that allow a guard to enter and exit the experiment, respectively. Agents are allowed to start their shift up to 30 minutes early and can exit a shift 20 minutes after the next shift is scheduled to arrive. Agents have additional activations for showing up for their shift when it starts and for leaving a shift when they are able. These activations drive the guards to typically show up on time and leave on time. In general, this means that guard shifts officially have a 20 minute overlap where guards can interact. In reality, guards sometimes arrived or left late. Modeling these transitions as actions allowed guards to arrive early, leave late, or not transition at all (stay in one place). In practice, the agent personalities modeled in the simulation tended to keep to the official shift times, plus or minus 10 minutes.

Table 6.4: Stanford Prison Experiment Transition Actions

Action	**Available To**	**Description**
Attempt Escape	Prisoners	Prisoner attempts to leave the experiment without permission, such as by a jailbreak. (Note: Since escape never occurred in the experiment, prisoners have a zero probability of actually escaping).
End Shift	Guards	If guard is on shift in the experiment, exit the experiment location.
Start Shift	Guards	If guard is off shift away from the experiment, enter the experiment location.

Table 6.5 lists the set of interpersonal actions, accompanied by a brief description of the meaning of the action within the simulation. This set of actions includes all of the frequency-recorded actions noted in Section 6.1.1 except for individuating and deindividuating references. The "referencing" actions were omitted because they were much more fine-grained than the other actions and would probably have significant overlap (i.e. threatening by name).

The interpersonal actions allow agents to interact with each other. While a majority of actions are negative or neutral, this does not necessarily mean that the majority of behavior would be negative. Positive interactions actions, such as helping, activate different parts of the GSP tree than actions such as insults and threats. This means that GSP trees are possible that would only perform neutral or positive actions. The actions chosen by each agent will depend heavily on its GSP personality weights. Additionally, the actions of an agent will depend greatly on the behavior of other agents. A guard that is generally passive may become abusive and sadistic when confronted with resistance, for example. The

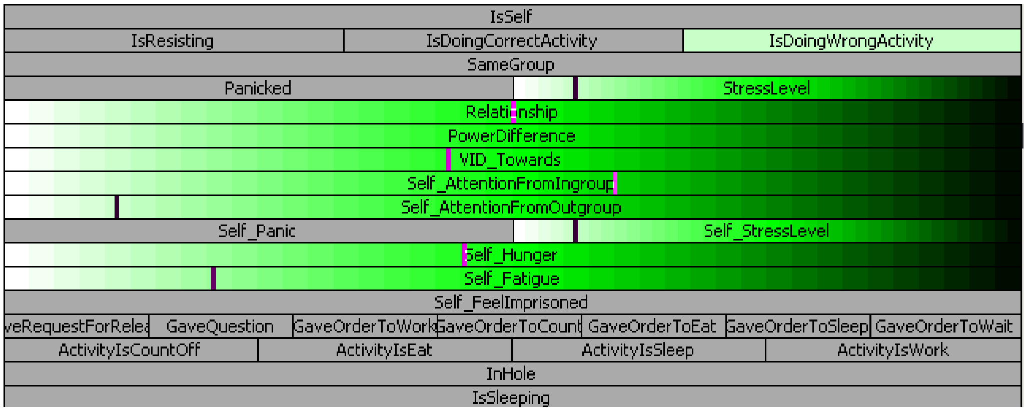
Table 6.5: Stanford Prison Experiment Interpersonal Actions

Action	**Available To**	**Description**
Command	Guards	Guard orders a prisoner to do the correct task, based on the schedule.
Demand Release	Prisoners	Prisoner demands to be let out of experiment.
Feel Imprisoned	Prisoners	Prisoner vocalizes that they cannot leave the experiment.
Help	All	Actor provides unsolicited help to target.
Information	All	Actor speaks to target, giving information about an event.
Insult	All	Actor calls target insulting names and/or describes them negatively.
Question	All	Actor requests information from target.
Physical Aggression	All	Actor physically handles target in a violent manner, such as an attack or a shove.
Remove From Hole	Guards	Guard removes prisoner from “The Hole” and returns them to regular activities.
Resist	Prisoners	Prisoner directly confronts guards, with the intention to change conditions for prisoners.
Threaten	All	Actor threatens target with negative consequences.
Throw In Hole	Guards	Guard initiates action to take prisoner to “The Hole,” a supply closet with a lock.
Use of Instruments	Guards	Guard threatens a prisoner by using a baton or other object as a symbol of authority.

context that modifies the base activations of each action is determined by a set of Perceptual Types (pTypes), as shown in Figure 6.2. The pTypes that define the activations for each action were calibrated as part of the initialization process, discussed in Section 6.1.3.

The activations produced by taking actions cause emotional responses for the actors taking those actions and for observers that view the events. In addition to the emotional effects, actions also cause direct effects. In the Stanford Prison Experiment simulation, there are four types of effects that occur: valence changes, authority changes, hunger changes, and fatigue changes. Fatigue and hunger are reduced by sleeping and eating, respectively. Valence and authority change as a result of interpersonal actions. Table 6.6 in Appendix H lists the direct effects of each action on valence and authority. Each column of the table represents the change in properties that occurs as a result of the action. For example, the table states that if an actor issues a command they gain authority while the target loses authority. The columns marked “Guards” indicate that taking a particular action changes the authority for all guards or the relationship of all guards toward the actor. Attempting escape and resistance undermine the guards as a group,

Figure 6.2: Stanford Agent PType Grid



so all guards in the experiment are affected by a prisoner taking these actions.

Table 6.6: Stanford Prison Actions - Valence and Authority Effects

	Authority			**Valence**		
Action	**Actor**	**Target**	**Guards**	**Actor → Target**	**Target → Actor**	**Guards**
Attempt Escape			-			-
Command	+	-				
Help					+	
Information					+	
Insult	+	-			-	
Physical Aggr.					-	
Threaten	+			-	-	
Release From Hole					+	
Resist	+		-			-
Throw In Hole	+	-			-	
Use of Instruments	+	-			-	

In addition to these effects, authority and valence are also changed due to prisoners obeying, disobeying, or resisting commands by guards. If a guard issues a command and a prisoner does not do the appropriate action, the guard's authority is reduced and the guard's valence toward that agent decreases. A prisoner resisting a guard results in a more severe decrease in both the relationship and the guard's authority. Alternatively, a prisoner obeying a guard improves the guard's valence toward them and slightly increases the guard's authority. Through this dynamic, de-facto leaders can emerge and the relationships between agents change as a result of behavior. As would be expected, valence tends to have positive feedback. Agents that initially take negative actions against each other tend to have deteriorating relationships, for example. This simple system of effects allows valence and authority to emerge from the interactions between agents, rather than assuming that agents would always develop the same relationships and power structure.

Simulation

The Stanford Prison experiment was simulated using a simultaneous simulation scheme. Under this simulation sequencing, all agents make their decisions and the results of all actions are executed simultaneously. After all actions have been executed, agents perceive all the events that have occurred simultaneously. This allows different events that have occurred to compete for attention. In this simulation, each time step represents 10 minutes, allowing agents to change their action every 10 minutes. The simulation ran for 693 steps, representing the time period from 4:30 PM on the first day until noon on the sixth day. This duration represents the approximate time period between when the prisoners were admitted to the experiment until the continuity of the experiment was permanently broken (no return to schedule).

Memes

Three memes of interest were studied: prisoner resistance (Resist), guards throwing prisoners in the hole (Throw In Hole), and feeling imprisoned (Feel Imprisoned). These actions were chosen for study as potential memes because they each showed signs of propagating through the groups over time, with clear early adopters. It should be stressed that these actions are only potential memes. The intent of simulation was to examine if treating these actions as memes better represents how these actions were expressed in the scenario.

The memes in the Stanford Prison Experiment scenario had two vectors of transmission. The primary vector was direct observation, where agents performed an action and other agents observed this and became aware of the affordance of that action. A secondary vector was possible by agents taking Information actions. The Information action for an agent contained information related to the first event returned by an agent's uncued recall from memory. If a meme action was recent and emotionally salient, an agent might learn about an affordance by talking with another agent.

Throw In Hole was chosen as a meme because it showed evidence of a clear early adopter: S_13 (John Wayne). S_20 was noted as imitating S_13 in some of the supporting materials. A document in the archives entitled "Remarks" asks, "Why did S_20 imitate John Wayne rather than S_15?" (real names have been replaced by coded numbers). It appears clear that some imitation in methods occurred among the guards. Finally, despite S_13 not arriving until the second shift in the experiment- the supply closet was not used as a solitary confinement until he first took this action. This gives a credible claim that other guards learned how to use the supply closet as “The Hole” through social learning. At the very least, it seems plausible that the guards moved from a general awareness of the potential for solitary confinement to it becoming an intuitively afforded action when punishing prisoners.

Resistance was chosen as a meme because it was studied explicitly within the experiment and S_05 was a clear resistance leader to start the experiment. While some prisoners enjoyed causing problems, S_05 resisted with the intention to change conditions in the experiment. This resulted in a general outbreak of resistance, which was eventually subdued. S_00, a late arrival, appeared to independently have an awareness of passive resistance, which he employed shortly after entering the prison. As such, if resistance was a meme, S_05 and S_00 appeared to be the original carriers.

Feel Imprisoned was chosen as a meme because it appeared to spread through the prisoner population as a result of S_05’s initial demand for release. After S_05 requested release and was convinced to stay, he reacted strongly and said that he felt he was really being imprisoned. Based on this meme spreading through the prison, other prisoners expressed that they felt imprisoned. While there has been debate in academic circles about if the experiment literally denied release to subjects, the logs show that many prisoners believed or feared that they would not be allowed to leave. One common theme was that they were in a prison, just a prison run by researchers rather than the government. Prisoners who expressed this meme did so with significant distress (ex. breakdown). Guards and prisoners both noted these strong reactions in their personal logs. The spread of this potential meme was inferred from the events in the hour by hour logs as well as excerpts from personal logs and letters written by prisoners. Rather than considering Feel Imprisoned as a meme with a static origin, it seems possible that this meme originated due to a prisoner feeling that they had been denied release. As a result, this meme could be spread to an agent by requesting release (and being denied) or by observing other agents expressing their feelings of imprisonment.

Studying these actions as memes is difficult because the hour-by-hour logs are not an exhaustive resource for the events in the scenario. Additionally, since only behavior can be observed, it is impossible to know when or if agents learn about affordances. For these reasons, the choice was made to study memes by examining the order that agents first took actions. As such, resistance would be studied by looking at the first agent that ever resisted, then the second, and so on. This method has the advantage that while not every instance of resistance would be recorded, the first time that an agent resisted the guards was notable and was noted in the hour by hour logs.

While this solves the original issue, a secondary problem is caused by the fact that in some cases more than one agent expressed a meme for the first time in a similar time period. Given that the hourly logs contain approximate times, it is difficult to state with certainty which agents expressed first in some cases. In general however, this was the exception rather than the rule. From the hour by hour logs and the supporting materials, the order that each agent first took each meme action was constructed.

Table 6.7: Stanford First Meme Expression Orderings

Throw In Hole	**Resist**	**Feel Imprisoned**
S_13	S_05	S_05
S_20	S_09	S_02, S_03
S_11	S_01, S_04	S_06
S_12, S_18	S_06	S_01
S_16,S_17,S_21	S_08	S_09
(S_15, S_19)	S_03	S_10
	S_00	S_00
	S_02	(S_04, S_08)
	(S_10)	

Table 6.7 lists the order that agents first took each action, as listed in the hour by hour log. Agents who share a row appeared to take the action in the same general time period, but with an unclear order. The final entry in each list contains agents for which no written evidence exists to show that these agents performed these actions. In most cases, the agents who were not noted to take certain actions make intuitive sense. S_15 and S_19 were noted as a "nice guard" and a "weak guard" respectively, and it is unclear if they ever initiated sending a prisoner to the hole. S_16 also appeared to be kinder to prisoners than other guards. Likewise, S_04 did not express a reaction to feeling imprisoned because he entered the experiment expecting that he would not be allowed to leave. This means that he would be considered a passive carrier, in some respects. These orderings were used as the ground truth, against which the simulation orderings were compared.

6.1.3 Stanford Prison Experiment Initialization

Before being able to simulate with the Stanford Prison Experiment scenario, the experiment had to be initialized with starting values for each model and activation tuning was necessary. Meme capable PMFServ agents needed to be initialized with starting values for physiology, authority, valences, and the weights for the GSP personality model.

Model Initialization

The physiology model contains tanks for the stomach (hunger), fatigue, and available caloric energy. No evidence from the experiment holdings suggested that the participants had significant differences in hunger or fatigue at the start of the experiment, so all agents were assigned the same initial values for each tank. All tanks were initially set to 80% full, to represent exertions related to gathering at the experiment site but having each agent be in good physiological condition at the start of the experiment.

Each agent was assigned as a member for their respective group. Prisoner group members were each assigned an authority of zero, the lowest possible authority. This represented their equal status as a group, as well as their low status within the experiment as a whole. Guard group members initially started with an authority of 0.25 (out of 1). This represented the authority conferred by their uniforms and role, while also representing initial equality between guards. Valences between agents were initially assigned to zero (neutral) for all relationships between different agents (on a scale from -1 to 1). Since an agent also has a valence toward itself, this valence was set to 0.8 (high valence) that assumed an initial state of positive self-attitude.

Initializing each agent's GSP model was a much more involved process than the other models, which were initialized based upon reasonable assumptions. The GSP personality models were initialized based upon the personality trait data from the Comrey Personality Inventory, the F-Scale, and the Mach test. The PMFServ GSP personality model can be set up to use these factors directly, but the modeling choice was made to utilize a pre-existing GSP tree structure. While harder to map the trait data into, this existing personality structure had performed well in PMFServ-based experiments such as Silverman and Bharathy (2005) and utilizes personality traits intended to correlate with behavior. Since the personality trait data was not used directly, a mapping between test concepts and GSP concepts had to be created.

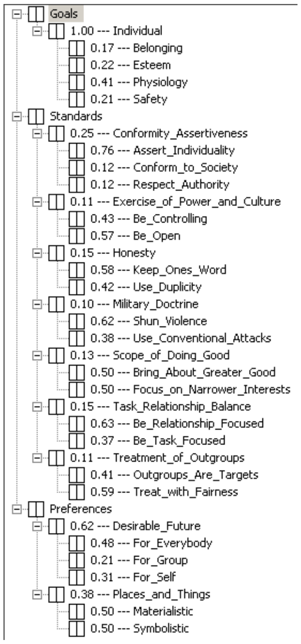
Table 6.3 shows the mapping between personality trait factors and GSP nodes. For reference, the GSP model of Guard S_13 is displayed to show the final product of the mapping algorithm. After each affected GSP node, the letter in parenthesis notes if the constructs are expected to have a high (H), moderate (M), or minor (L) correlation. Using this map, an algorithm generated GSP tree weights based upon the raw values for each personality trait metric. The algorithm is contained in Appendix H. The algorithm accepts normalized measures (between 0 and 1) and returns an appropriately weighted GSP tree. Since the raw trait data is not normalized, each measure was renormalized to fit a range between 0 and 1.

The attention model also had a pair of parameters that had to be set up as part of the model initialization. As noted, the attention model has parameters that set the maximum number of attended events and an inattention salience

Figure 6.3: Map: Measured Traits to GSP Nodes

Measured Trait	Related GSP Nodes
Trustworthiness	Keep_Ones_Word (H)
Orderliness	Be_Controlling (H)
Conformity	Conform_To_Society (H), Respect_Authority (M)
Activity	Physiology (H)
Stability	Assert_Individuality (M)
Extroversion	Belonging (M), Esteem (M), Be_Relationship_Focused (L)
Masculinity	Use_Conventional_Attacks (H), Conform_To_Society (M), Belonging (M)
Empathy	Be_Relationship_Focused (H), For_Everybody (H)
Machiavellianism	Be_Task_Focused (H)
F-Scale Value	Outgroups_Are_Targets (H), For_Group (H)

Figure 6.4: S_13 GSP



```
Goals
  1.00 --- Individual
    0.17 --- Belonging
    0.22 --- Esteem
    0.41 --- Physiology
    0.21 --- Safety
Standards
  0.25 --- Conformity_Assertiveness
    0.76 --- Assert_Individuality
    0.12 --- Conform_to_Society
    0.12 --- Respect_Authority
  0.11 --- Exercise_of_Power_and_Culture
    0.43 --- Be_Controlling
    0.57 --- Be_Open
  0.15 --- Honesty
    0.58 --- Keep_Ones_Word
    0.42 --- Use_Duplicity
  0.10 --- Military_Doctrine
    0.62 --- Shun_Violence
    0.38 --- Use_Conventional_Attacks
  0.13 --- Scope_of_Doing_Good
    0.50 --- Bring_About_Greater_Good
    0.50 --- Focus_on_Narrower_Interests
  0.15 --- Task_Relationship_Balance
    0.63 --- Be_Relationship_Focused
    0.37 --- Be_Task_Focused
  0.11 --- Treatment_of_Outgroups
    0.41 --- Outgroups_Are_Targets
    0.59 --- Treat_with_Fairness
Preferences
  0.62 --- Desirable_Future
    0.48 --- For_Everybody
    0.21 --- For_Group
    0.31 --- For_Self
  0.38 --- Places_and_Things
    0.50 --- Materialistic
    0.50 --- Symbolistic
```

that affects the likelihood of paying attention to events that are not modeled. The maximum number of attended events was set to 4, based upon research that shows that typically humans only have sufficient working memory to keep track of 4 items at once (Cowan, 2001). The inattention salience was set to a low value, 0.28. This value was selected so that a single, maximally salient event would be attended with a 92% probability. This value was chosen based on the maximum recall rate for across the experiments used to set the salience components in Section 5.2.5.

Calibrating Activations

Activations for actions were calibrated manually, since insufficient data existed to automate a training algorithm that would produce meaningful activations. While various methods of automated calibration were considered, a machine learning approach was deemed too likely to produce unintuitive semantics. Agents might take the actions at the right frequencies, but for the wrong reasons. The combination of limited training data and the richness of personality data made this outcome almost unavoidable. As a result, the activations were calibrated manually to ensure face validity was maintained.

For manual calibration, the first values for activations were set ad-hoc based upon the best guess. This stage was not intended to get precise values, but identified contexts where certain GSP nodes that should be activated. For each pType-action pair, the set of possible GSP node activations were selected.

The second step of calibration involved balancing different action contexts, using a fully normalized GSP tree. In a fully normalized GSP tree, all nodes are valued equally. This makes calibration easier, since the subjective utility of any action reduces to a sum of linear activation values. This calibration method generated rankings of the same action under different contexts, as well as rankings of different actions under similar contexts. This allowed tuning activations for the same action under different contexts, to establish which contexts were preferred for performing that action. It was also used to set conditions where one action would strategically dominate a different action (Shoham & Leyton-Brown, 2009). This tuning phrase established reasonable values for activations under each of the possible pType contexts for an action.

Finally, activations were calibrated by simulating the first 20 hours of the experiment repeatedly and calculating the frequency that actions occurred during each scheduled activity. Calibration was performed using the Fully Known experimental case which has all agents know all actions (described in the following section). The intention of this calibration phase was to ensure that actions occurred with the appropriate relative frequencies with respect to each other. The tuning script generated a report listing the frequency that each action occurred. This report was compared against the expected frequency for each type of action, if known.

Table 6.8: Activation Training Frequencies

Action	**Time Period**	**Frequency (Events/10 min)**
Command	Count Off, Eat, Work, None	6.46
Help	Count Off, Eat, Work, None	(only 1 recorded)
Information	Count Off, Eat, Work, None	3.02
Insult	Count Off, Eat, Work, None	3.10
Insult (by S_13)	S_13 on Shift	1.55
Resist	Count Off, Eat, Work, None	1.58
Threat	Count Off, Eat, Work, None	1.48
Use Of Instruments	Count Off, Eat, Work, None	1.34

Table 6.8 notes the frequencies used for training the relative frequencies of actions. These frequencies could not be used for direct comparison, since the simulated agents only generate one event per 10 minute interval. Tuning to these frequencies would be impossible, since the total number of events would be higher than agents could generate. Instead, the metrics were used to examine the relative differences between the expected count of each action as compared with the actual count. While the simulated frequency of all actions was lower than the ground truth, the calibration goal was to ensure that each action occurred in a similar proportion compared to each other.

A second set of metrics measured the baseline actions during each period. Based on the archival materials, prisoner agents spent a majority of their time performing scheduled activities rather than engaging in interpersonal actions. The training metrics assumed a high frequency for a scheduled action during its assigned time period. It was assumed that prisoners performed the scheduled activity 75% of the time during count off periods, 80% of the time during eating periods, 90% of the time during sleep periods, and 75% of the time during work periods. While these activities were not specifically measured during the experiment, it was assumed that prisoners performed their assigned activities during periods of no incidents.

Using these metrics, the activations were trained to approximate the expected distributions over the start of the experiment. This allowed the simulation to better match the expected behavior from the actual experiment. However, this calibration was not intended to ensure that the behavior of the simulation exactly matched the ground truth frequencies. Due to random and chaotic elements in the simulation, individual simulation runs can deviate from these distributions. However, it provided a useful tool for ensuring that agents would use each of the modeled actions in reasonable circumstances and at reasonable rates.

6.1.4 Stanford Prison Experimental Cases

Three experimental cases were designed for the Stanford Prison Experiment, each intended to represent one hypothesis for the origin of the potential memes within the experiment. The three experimental cases will be referred to as Full Knowledge, Authority condition, and Hypothesis condition. These cases differed based on the agents who were initially aware of the affordances for Throw In Hole and Resistance. Feel Imprisoned was allowed to emerge under the same conditions across all runs, as a result of a prisoner being denied release. The Full Knowledge case assumed that no meme reproduction occurred because the agents were aware of all affordances at the start of the experiment. This would mean that even if memes existed, they would be at saturation and agents would not learn new affordances socially.

The Authority condition is based upon the hypothesis that guard cruelty was fostered due to information presented to participants during their orientation (Reicher & Haslam, 2006). This condition assumes that participants were presented an example of each meme at the start of the experiment, as part of an orientation. For example, the guard orientation might have included a demonstration of how to throw a prisoner in “The Hole.” This condition presents each agent with an event that shows the Experimenter agent taking the meme action, for each meme action. This condition will result in a random subset of agents receiving each meme, with the bulk of attentional salience driven by the authority of the Experimenter agent. Agents will differ in their reception of the meme based on if they can repeat it (transferability) and if they have a similar personality (similarity).

The Hypothesis condition assumed that certain agents acted as seeds for the meme to spread through the population. This is referred to as the Hypothesis, as it is intended to test if memes are a plausible mechanism for explaining the order that memes were expressed. In this condition, S_13 was the only agent initially aware of the Throw In Hole affordance when perceiving other agents. Correspondingly, the Resist action was only afforded to S_05 and S_00 at the start of the experiment. Since S_00 did not enter the experiment until the fifth day, he could only have passed along the meme to S_02 or S_10 for the first time. All other agents had already expressed resistance by that time.

If the Meme Origin condition shows a better match to the first expression ordering, this would imply that memes might have had a role in transmitting certain affordances through the Stanford Prison Experiment subjects. This would not necessarily imply that the exact memes stated would be the memes involved, however. As explored in Section 4.2 and Appendix B, other forms of social learning can have similar outcomes to socially learned affordances.

Even if this order was influenced by the spread of memes, this does not prove that the meme was a socially-learned affordance. With that said, these actions are sufficiently complex and specific to the experimental condition that it is reasonable to suggest that their affordances might have been learned through the course of the experiment. If simulating memes improves models this order well, this will only imply that some memes affected the order that these actions were expressed. The exact nature of any such memes cannot be known, since the original Stanford Prison Experiment did not attempt to measure any such learning in detail.

The Stanford Prison Experiment scenario was simulated for 30 runs under each experimental condition, for a total of 90 runs of the Stanford Prison Experiment. This provided sufficient data to apply a variety of analyses, including first expression ordering and diffusion rate estimation. The data collected from this experiment was used for internal validation measures, as well as externally supported metrics such as the first-expression of meme actions.

6.2 Scenario 2: Iraqi Village

The second scenario modeled was Hamariyah, an archetypal Iraqi village based on a human terrain data set. This scenario was generated by the ACASA lab, utilizing data provided by the US Marine Corps (USMC) (Silverman, 2010). This scenario stresses the model of memes, connecting it with day to day economics and a rich society based on human terrain data. Compared to the Stanford simulation, the Iraqi village reduced agent attention capabilities of events, used a longer time step value, and used behaviors that unfold over time rather than occurring immediately within a time step. This scenario was designed to evidence completeness, that the model can be used meaningfully in different context. It also uses competing memes, to allow examination of selection effects. While the Stanford Prison Experiment had separate memes for each group, the Iraqi Village allows its memes to be reproduced by any agent. Two memes exist in this scenario: giving information to the US-backed government and planting an IED by a government building. Since this framework had pre-existing actions, these memes competed against each other and against the existing action set, which primarily models daily life.

6.2.1 Iraqi Village Data Sources

The Hamariyah Iraqi village is a fictional village created by the USMC for the purposes of training. While it is not modeled on any specific Iraqi village, it integrates common social structures, cultural elements, and personality trait profiles that would be representative of an Iraqi village. The original data used to design this village is described in Silverman (2010), stating:

> The USMC folks from 29 Palms generated Hamariyah and descriptions of the town history, its 200 residents, 3 tribal groups, families, jobs, institutions, inter-factional grievances, and so on. This is a paper-based description, though some of it was provided in comma separated value (csv) files that we recast into spreadsheet workbooks that were then read by the PMFserv model constructor. (p. 25)

The Hamariyah village was used as the base scenario for simulation, with some modifications. The village itself was built off of human terrain data contained in comma separated value (CSV) files. These files outlined the set of agents in the scenario, with information about each agent's personality and social position. Table 6.9 lists relevant data available for each agent, which was used to generate the PMFServ scenario. In addition to these data fields, written materials described the village's members, backstory, and the relative advantages of each group.

Based on this information, modelers and programmers populated the Hamariyah NonKin village scenario- which has been under development from approximately 2007. In addition to the agent properties, information about typical daily life actions and insurgent activities were present in the data sources used to design the village.

Table 6.9: Hamariyah Agent Data Fields

Agent Property	Description
Age	Age of the agent, as an integer
Attitude Toward MNF	Attitude toward multi-national forces (US), in terms of Pro-MNF, Anti-MNF, and Neutral
Employment Level	If an agent is employed full-time, part-time, or unemployed
Ethnicity	Ethnic group, such as Arab or Kurd
Family Name	Name of the family the agent belongs to
Gender	Agent gender, either male or female
Internal Group	A subgroup an agent belongs to, if applicable.
Internal Role	Role of the agent in their main group, chosen from Leader, Core Follower, and Fringe Follower.
Kinetic Special Skill	Special knowledge agent has about weaponry (i.e. placing IED)
Marital Status	Relationship status: Married, Single, Widowed
Military Experience	Level of military training the agent has
Occupation	The agent's job in the scenario
Personality Archetype (GSP)	The type of personality for the agent, selected from a list of 12 personality models. A GSP personality model was estimated to represent each personality type.
Religion	The religion of an agent, chosen from Shia, Sunni, Christian, or None.
Tribe Name	The tribal group an agent belongs to, chosen from Heremat, Shumar, and Yousif.

6.2.2 Iraqi Village Scenario Design

Given the scope of the NonKin Village project, it is infeasible to explore every aspect in detail. Instead, the following sections will highlight the key scenario features and note any modifications that were necessary in order to study memes within village. For more detailed information about the scenario Silverman (2010) discusses the design history of the Hamariyah village, while Silverman et al. (2009) overviews the general NonKin village functionality. Since the NonKin village was being used as a simulator rather than a game, certain elements of the framework noted in Silverman et al. (2009) were not utilized. In particular, 3-D representation and detailed conversations were disabled because no user agents existed to interact with the village. Otherwise, this scenario may be considered an extension of the existing NonKin village scenario template.

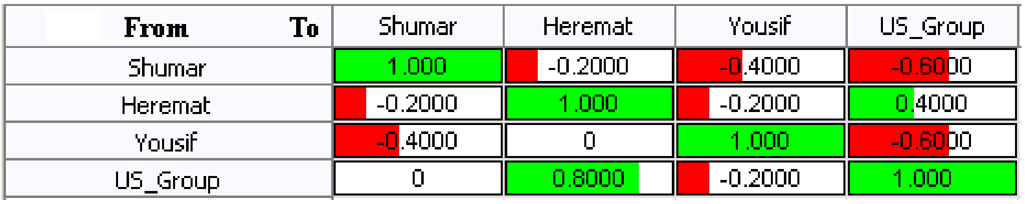
Entities

A NonKin village contains three types of entities: groups, agents, and structures. These entities differ from the Stanford scenario in a few notable ways. Firstly, all agents utilize a slightly different model set. While all the new meme-enabling models are the same, NonKin agents utilize modified decision, perception, and physiology models. These models work similarly to the standard PMFServ models, but are modified to allow agents to be asynchronously driven- such as by interacting with a 3D engine. A second major difference is that in the NonKin village, the location of an agent is very important. Certain actions are only available at particular locations, such as sleeping at home or working at their workplace. This makes buildings and other structures important entities in the NonKin village.

The Hamariyah scenario contains 200 agents, split into three main groups: Heremat, Shumar, and Yousif. In addition to the agent groups, a US group was present in the scenario to allow group relationships and ownership of buildings in the village. The original scenario file split some of these groups into subgroups, such as particular militia cells. Since these groups included only a handful of agents, all subgroups were collapsed back into the three primary groups. Agents that had leadership roles in subgroups were assigned to the main group as followers with a high level of authority. In addition to agents being members of groups, structures in the NonKin village are tagged by their group affiliation. This allows agents to see if buildings belong to their group, a group they like, or an unfriendly group.

These relationships are determined by the group to group valences, as shown in Figure 6.5. The Heremat group is generally friendly to the US and controls the local police force, but is not a very big group. The Shumar group is a primarily Sunni group unfriendly toward all other groups, especially the US Group. It is the largest group, with a majority of its members working as merchants or tradesmen. The Heremat and Shumar groups both have members working as part of the local government. The Yousif group is a primarily Shia group, with higher than 60% unemployment and religious leaders in higher positions of authority.

Figure 6.5: Hamariyah Group Valences



From To	Shumar	Heremat	Yousif	US_Group
Shumar	1.000	-0.2000	-0.4000	-0.6000
Heremat	-0.2000	1.000	-0.2000	0.4000
Yousif	-0.4000	0	1.000	-0.6000
US_Group	0	0.8000	-0.2000	1.000

The agents in the Hamariyah scenario include some agents that are intended for simulation and other agents that are intended for scripted actions, such as external combatants entering the area. To keep the village streamlined and populated with agents with full sets of human terrain data, only a subset of agents from the full Hamariyah village were used for simulation. Of the total set of agents, 72 agents were simulated for meme analysis: 11 Heremat members, 38 Shumar members, and 23 Yousif members. These members utilize 10 separate GSP personality models, as shown in Table 6.10.

Table 6.10: Number of Agents Using Each GSP Model, By Group

GSP Model Description	**Heremat**	**Shumar**	**Yousif**
Al-Qaeda Iraq (AQI)		2	
Baathist	2	8	13
Child		1	
Shia Imam			3
Iraqi (Sunni)		10	
Mayor/Official	2	1	
Merchant		9	
Policeman	2		1
Tea Man (Tea house owner)	3		2
Woman	2	7	4

While it would be impossible to describe the full modeling process of each of these GSP models, they do have some notable differences. The AQI and Baathist GSP models are more accepting of violence and asymmetric tactics than other personalities. They also tend to be task focused, rather than relationship focused. Conversely, Police GSP models accept violence but prefer to use conventional tactics and be relationship focused (i.e. negotiate). The Imam and Mayor/Official GSPs represent leaders and potential leaders. These agents focus on asserting individuality and leadership goals, rather than primarily focusing only on day to day goals. Merchants, women, and children place a low value on violence compared to other personalities. Across all personalities, agents value outcomes that benefit themselves or their group but place little value on outcomes that benefit other groups. These personality weights affect how each agent will respond to the available memes: giving information and planting an IED.

The structures in NonKin village are important elements of the environment. Three structures play a major part in each agent's life: homes, workplaces, and mosques. Many actions are only available in particular locations. Each agent has a home in the simulation, where they perform actions such as sleeping, eating, and socializing. Workplaces fill a double role as places of employment and places of business. Employed agents have a workplace where they typically go during their work shifts in order to earn money. These same workplaces provide services such as selling food. Mosques are a special form of workplace where agents may come to pray. Agents consider the group affiliation of businesses before patronizing them, as well as the religious affiliation of a house of worship before attending. A significant portion of each agent's day will be traveling to various buildings and performing daily life tasks at these locations.

A special structure exists to be the target of meme actions. This structure is affiliated with the US group and is named the "Government Meme Target." This structure does not have any workers modeled and only allows agents to visit the building, give information, or plant an IED near the building. As the name implies, this structure exists in order to allow agents to perform meme actions on an object representing US interests in the region.

Actions

Including memes, the Iraqi village PMFServ file contains 57 different actions which can be taken by agents in the scenario. These actions range from complex multi-stage actions (i.e. go to market and buy food) down to niche actions for forcing entry into a building. The actions in the Iraqi village were not modified in any way except by the addition of the two new meme actions. The most common actions agents take within Hamariyah village are those related to daily life. These actions include moving from one building to another, entering/exiting buildings, buying food, working, socializing, praying, sleeping. Agents are also able to take less common actions such as attacks, shootings, and hiring/firing employees but these actions are infrequent. Based on the actions that agents most commonly perform, agents will commonly be deciding between going about a normal day or taking an extreme action for or against the US.

Simulation

The Iraqi village runs in 30 minute steps, such that every agent is allowed to pick a new action every 30 minutes. Each simulation run lasted for 2 days in simulation time (48 decisions per agent). During the 30 minute interval, agents travel between locations and perform components of the actions they committed to at the start of their decision cycle. Unlike the Stanford scenario, the Iraqi village is simulated asynchronously. This means that each agent takes their action separately, rather than at the same time. From the standpoint of attention, this means that agents will only observe a maximum of one event at any particular time. This means that events compete against inattention salience rather than each other. Otherwise, the Hamariyah scenario simulation runs similarly to the Stanford scenario.

Memes

The memes modeled in the Hamariyah Iraqi village were Give Information and Plant IED. Both of these memes could only be performed on the "Government Meme Target" structure. The Give Information action represents acting as an informant to the US. Giving Information is a risky action in this context, because anti-US forces could try to eliminate local informants. The meme for Give Information is the learned affordance from an agent can go to the US structure to inform on dangerous members in the village. The Plant IED action is an opposite and competing action. This action involves being willing to take an IED from local militia groups and place it in the vicinity of the US structure. As with Give Information, this action has inherent risks that give it negative activations for personal safety.

For both memes, most agents would need to feel strongly about supporting or opposing the US Group in order to take these actions. However, these memes have other appeals. The Plant IED action can appeal to agents that greatly value violence and attacks, for instance. Similarly, the Give Information action can appeal to agents that support building relationships and negotiation. These characteristics are part of what determines which agents will be likely to express these memes within the village.

Due to limitations on simulation length, both memes have been made more attractive than they would be in the real world. This allows for better examination of relative expression rates and diffusion, since it avoids runs where no agents express the meme. Due to this modeling choice, the village simulation will not be a good a predictor of agents that would never express the meme. However, it increases the ability of the simulation to work as a relative predictor of agent's preferred meme. Since this simulation is intended to examine meme selection, this is beneficial for the analysis. However, it does mean that the total number of people who express either meme will be higher than one would expect in a real world scenario- especially given the time frame.

6.2.3 Iraqi Village Initialization

The Hamariyah village was a pre-existing scenario, so minimal initialization was necessary. Only the new models required initialization, such as the attention model. The attention model, as previously noted in Section 6.1.3, has values to set the maximum simultaneously attended events and the inattention salience. The Hamariyah simulation works asynchronously so agents will be presented with at most one event at any given time. This meant that the maximum number of attended events was set to 1. The inattention salience also had to be significantly increased compared to the Stanford Prison scenario. Since the Stanford Prison was in a controlled environment, distractions were minimal. In a real village situation, intermittent distractions would be more prevalent. Presenting events one at a time also requires an additional inattention salience, to account for the lack of competition that would otherwise exist. Based on these factors, the inattention salience level was set to 8.0. This meant that an event of typical salience (about 1.6) would have a 1 in 6 chance of being attended. While this may seem low, with 72 NonKin agents interacting in a small village, this means that an agent observing all of these events would be expected to observe 12 events. This was the only model that required initialization for this simulation.

6.2.4 Iraqi Village Experimental Cases

The Hamariyah scenario utilized two experimental cases: a Hypothesis case and a Randomized case. The Hypothesis case assumed that a particular set of 6 agents initially knew each meme, based upon their roles in society. In the hypothesis case, Give Information was initially known by HAM003, HAM004, HAM005, HAM0021, HAM041, and HAM084. These agents were chosen because they were members of the local police (HAM003, HAM004, HAM005, HAM0021) or involved with the local government (HAM041, HAM084). Agents in the police force and government could be expected to be aware of how and where to provide intelligence to the US forces in their area. Plant IED was initially known by agents HAM059, HAM060, HAM075, HAM081, HAM120, and HAM130. These agents were all categorized as anti-US and their Kinetic Special Skills listed them as a “IED Maker” or “IED Emplacer.” These agents were specifically noted in the human terrain data as having IED skills, so they started with the affordance to Plant IED in the scenario. This scenario was intended to represent the transmission of competing memes under realistic conditions.

The Randomized case took the opposite approach. At the start of each run, 6 agents were randomly chosen to start with the Give Information affordance and another 6 agents were randomly chosen to start with the Plant IED meme. No constraints were placed to allow an agent to start with only one meme, so it was probabilistically possible for one agent to start with both memes. This scenario was intended to examine the patterns of meme transmission that exist when memes are available to agents that would not normally be expected to carry them. This scenario allows examining scenarios such as passive carriers, agents that start with the meme but never express it.

Twenty runs were simulated under each experimental condition, with data collected from each run for analysis. The Hypothesis condition always started with the exact same initial conditions, while the Random condition started with a different random set of agents aware of each meme on each run. These experiments provided interesting results and insights into competition of memes in a rich multi-agent environment.

Chapter 7

Analysis and Results

The same set of metrics were recorded for the Stanford simulation and the Iraqi village simulation. For each simulation run, the simulation system recorded the actions, emotions, attention focus, and learning for each agent. The analysis of this data attempts to examine how meme transmission occurs at the individual level and how meme diffusion occurs at the societal level. The types of analysis applied can be split into three categories: internal validity, external validity, and exploratory analysis. This section will begin by discussing the data collected and the paradigms used to analyze this data. Internal validity measures are discussed next, examining internal validity measures from both scenarios. Following this, an analysis of the Stanford simulation will present external validity measures and an exploratory analysis of meme transmission trends. Finally, an exploratory analysis of the Hamariyah Iraqi village will show trends within this larger and more diverse simulation.

7.1 Simulation Data Collected

During each simulation run, data was collected after each time step. This data was logged into four data tables: simulation events, agent emotions, agent meme awareness, and affordance transmission. The simulation events table logged every action that occurred during a simulation run, coded by the simulation time. Table 7.1 displays the data collected about simulation events. This data table logs to the standard data present in PMFServ events. This data was collected to examine action frequencies and to examine when memes where expressed.

The agent emotion table captures each agent's set of emotions at each time step, which are the result of their actions and the actions of other agents. Table 7.2 displays the data collected about agent emotions during simulation.

Table 7.1: Simulation Events Data Table

Data Field	Description
Simulation Step	Time that the action occurred
Actor	Name of agent taking the action
Action	Name of action agent initiated
Target	Name of target of the action (if targeted)
Result	Result of the action

Joy, Distress, Pride, Shame, Liking, Disliking, Gratification, and Remorse are emotions generated by the PMFServ emotion model, which is based on the Ortony et al. (1988) formalization of emotions. Each of these emotions has a range between 0 and 1, with 0 being none of that emotion present and 1 being the strongest feeling of that emotion. Multiple emotions can be present simultaneously under this system. These are discussed in detail in other papers, as they are core parameters of PMFServ (Silverman et al., 2006).

Table 7.2: Emotions Data Table

Data Field	Description
Simulation Step	Time that the emotions were measured
Agent	Name of agent who has these emotions
Group	Name of the group that the agent belongs to, if any
IsOnShift	True if the agent is on shift (present in experiment), else False (Stanford Experiment Only)
Joy	Joy of an agent due to short term goal successes
Distress	Distress of an agent due to short term goal failures
Pride	Pride of an agent due to success following personal standards
Shame	Shame due to failures in following personal standards
Liking	Like of the world state, based on long term preferences
Disliking	Dislike of the world state, based on long term preferences
Gratification	Positive compound emotion that combines joy and pride
Remorse	Negative compound emotion that combines distress and shame
Aggregated	Emotion term representing the total emotional state
Stress	Integrated stress of the agent (from the stress model)

The Aggregated term was calculated during data collection and is a sum of those eight emotions, where good emotions are taken as positive and bad emotions are taken as negative. This sum is divided by 4 to fit between -1 (dysphoric) and 1 (euphoric) and can be thought of as an estimate of the valence of an agent's current emotional state. Equation 7.1 shows how the Aggregate emotion value is calculated, based upon an agent's other emotions. The Stress term is calculated by a separate PMFServ model, which calculates an integrated stress value that is based upon emotional stress, time pressure, and fatigue (Silverman et al., 2006). This term varies between 0 and 1, where 0 is completely unstressed (nearly unconscious) and 1 is a state of panic. This data table provides a summary report of the agents' affective states over time. This data was collected to compare the emotional trends in the Stanford Experiment with those reported in Haney et al. (1973a).

$$\begin{aligned} AggregatedEmotion = &\frac{1}{4}((Joy - Distress) + (Pride - Shame) + \\ &(Liking - Disliking) + (Gratification - Remorse)) \end{aligned} \tag{7.1}$$

Agent meme awareness is a table that is generated from the each agent's memory model. This table is formatted as shown in 7.3. This table has a column for each meme measured, to monitor when each agent became aware of each meme. This table is a simplified probe of the memory model, which is either knows a meme (familiar) or doesn't know about a meme. This data was recorded to help measure agent learning.

Table 7.3: Meme Awareness Data

Data Field	**Description**
Simulation Step	Time that agent knowledge was probed
Agent	Name of agent whose memory was checked
(Meme Name)	True if agent is aware of an affordance used as a meme, else False
(...)	(Additional fields for other memes)

The affordance transmission data records if an agent attended and learned from each event that occurred during the scenario. This data was recorded for meme-related events and unrelated events, since the attention step is inherently competitive. This recorded N^2 entries per step, where N was the number of agents in the scenario- making the data table very large (approximately 250,000 entries per run for the Stanford Prison simulation). Table 7.4 displays the fields recorded within the transmission data table. This table notes three possible levels of processing an event: "Can Observe," "Attended," and "Learned." Each stage requires the prior stage to be true. An agent must physically be able to detect an event to attend it and must attend an event to learn from it. In this way, physical meme barriers are differentiated from attentional issues and learning issues. It also stores the factors that contribute to attention salience, such as novelty, motivated attention, and other event salience components noted in 5.7. One focus of analysis will be the "Attended" parameter, as mediated by these salience components. Additionally, this interaction data contains information about which agents pay attention to which other agents- useful for examining the emergent social network for attention.

Table 7.4: Affordance Transmission Data

Data Field	Description
Simulation Step	Time that agent knowledge was probed
Observing Agent	Name of agent examining an event
Event Acting Agent	Name of agent performing initiating event examined
Event Action	Name the action initiated by the acting agent
Can Observe	If true, the observing agent can physically detect the event (i.e. close enough to see)
Attended	If true, the observing agent attended this event
Learned	If true, the observing agent recorded the action from this event
Total Salience	The total attentional salience of the event, as a weighted sum of attention salience factors
Authority	Authority of the actor of the event
Conformity	Conformity due to number of agents engaged in the event action at this time
InGroup	If observer and actor share the same primary group, this is 1, else 0
Motivation	Motivation to gain the outcomes of the event
Novelty	Novelty factor of the event
Reference Group	Amount that an observer uses the actors group for social cues
Selection	Amount of active attention to the actor of the event
Similarity	Similarity of personalities between observer and actor
Transferability	If observer can take event action, this is 1, else 0
Valence	Amount that observer likes actor (0 is disliked, 1 is liked)

7.2 Analysis Methodologies and Techniques

A variety of analytical techniques were applied to examine relationships in the simulation data collected and to compare these against external holdout data. Each of these methods will be discussed briefly in this section, to allow a more coherent discussion of the results in the following sections. To assist in batch analysis of the data, all computational analyses were performed using Python code, pre-existing Python packages, or other statistical packages wrapped in Python (ex. the R stats framework).

7.2.1 Correlation Analysis

Correlation matrices were generated from the transmission data in order to examine the effects of social and situational factors on attention and learning.

This correlation analysis provides information about the strength of each of the input variables (ex. authority) on the dependent variable: learning from an event. These correlations can be compared against the empirical findings from the studies used to design the cognitive components, as a test of internal validity. Additionally, the correlation analysis provides information about relationships between factors. While from the attention model's standpoint, each factor is independent- their values may be influenced by common factors. Correlation analysis was completed using the SAS 9.2 software, generating both the Pearson and Kendall correlation matrices.

7.2.2 Multivariate Regression

Multi-variate generalized linear regression techniques will be used for internal validation testing of the attention process for agents. Internal validation is important, since the PMFServ agent cognitive model contains dozens of interacting cognitive components. In particular, the attention salience component integrates the input from ten other new cognitive components. Additionally, the salience term must itself be used by the attention model to determine the probability that an event is attended. While each component was carefully designed and tested, internal validation of the agents' attentional responses was an important check to make sure that all components were implemented as intended. The internal validity test design was based upon the Affordance Transmission Data, described in Table 7.4. Each row in this data table contains the values of the inputs to attentional salience (e.g. the novelty of the event), accompanied by whether or not the agent paid attention to that event. The regression was intended to validate that the inputs to attention had the correct relative importance and that their signs were correct.

A multi-variate linear regression assumes that the data is a set of N observations in the form $y_i = \overrightarrow{X_i} \cdot \overrightarrow{\beta} + \epsilon$ for each observation $i \in \{1, ...N\}$. In this formulation y_i is a response variable, $\overrightarrow{X_i}$ is a vector of inputs, $\overrightarrow{\beta}$ is a vector with a weights for each input, and ϵ represents unexplained error. The regression algorithm attempts to estimate the weights ($\overrightarrow{\beta}$) that best explain the response variables as a function of the input variables.

Regression techniques were used to examine how the simulated agents oriented their attention to events, which in turn controls the events they learn from. As noted in Section 5.2.5, the events that agents pay attention to are probabilistically selected as a function of the attentional salience of each event. The attentional salience term was implemented in a PMFServ cognitive component as a function in the form shown in Equation 7.2 (a copy of Equation 5.9 in Section 5.2.5). Each of the weight terms (w_i) was initialized with a "best guess" value from examining associated literature.

$$\begin{aligned} Salience(e) = & w_0 \cdot \text{Novelty(e)} + w_1 \cdot \text{MotivatedAttention(e)} + \\ & w_2 \cdot \text{SelectiveAttention(e)} + w_3 \cdot \text{Transferability(e)} + \\ & w_4 \cdot \text{Authority(e)} + w_5 \cdot \text{Conformity(e)} + w_6 \cdot \text{Similarity(e)} + \\ & w_7 \cdot \text{Valence(e)} + w_8 \cdot \text{InGroup(e)} + w_9 \cdot \text{ReferenceGroup(e)} \end{aligned} \tag{7.2}$$

Logistic regressions were used to estimate these factors from the model when applied to the Attention and Learning data collected. This was done by finding and examining the regression β coefficients, based upon a data set of agents examining events. The regression formula used to estimate the coefficients is shown in Figure 7.3. This formula mirrors the one for the attention salience equation, but with three differences. Firstly, an intercept term m was estimated by the regression (as is typical in a regression). Secondly, the response variable is a binary output designating if the event e was attended or not. Finally, there is error term ϵ because this is a regression equation form- this represents any unexplained variance.

$$\begin{aligned} Attended(e) = & \beta_0 \cdot \text{Novelty(e)} + \beta_1 \cdot \text{MotivatedAttention(e)} + \\ & \beta_2 \cdot \text{SelectiveAttention(e)} + \beta_3 \cdot \text{Transferability(e)} + \\ & \beta_4 \cdot \text{Authority(e)} + \beta_5 \cdot \text{Conformity(e)} + \beta_6 \cdot \text{Similarity(e)} + \\ & \beta_7 \cdot \text{Valence(e)} + \beta_8 \cdot \text{InGroup(e)} + \beta_9 \cdot \text{ReferenceGroup(e)} + m + \epsilon \end{aligned} \tag{7.3}$$

This regression was useful for testing the implementation of the total attention system. Since attention to events is a function of their salience, the sign and importance of each input into salience should match its sign and importance to attention in the cognitive model. Equation 7.4 restates the relationship between attention and attentional salience (explained in detail in Section 5.2.5). In this equation, E is the set of all simultaneously observable events, E_{Att} is the set of already attended events, s_e is the salience of an individual event e, and s_I is the inattention salience.

$$P[e = Attended] = \begin{cases} \frac{s_e}{s_I + \sum_{e \in E \setminus E_{Att}} s_e} & \text{if } e \in (E \setminus E_{Att}) \\ \frac{s_I}{s_I + \sum_{e \in E} s_e} & \text{No Event Attended} \\ 0 & \text{if } e \in E_{Att} \end{cases} \tag{7.4}$$

When only one single event is presented at a time, Equation 7.4 reduces to Equation 7.5. This means that for a single event, the probability that an event is attended is quite similar to the logistic function. This indicates that in this simple case, the regression should give a good estimate of the attentional salience function so long as the inputs are all independent.

$$P[e = Attended] = \frac{s_e}{s_I + s_e} \tag{7.5}$$

Three such data sets were examined using this approach. The first data set was generated by artificially generating sample events. These events were specifically created with each salience input was independent and selected randomly from its possible range. A single agent processed these events and returned if that agent paid attention to the event or not, populating a data set in the form described in Table 7.4. Attention in this system follows the form shown in Equation 7.5.

The other data sets examined were collected from the Stanford Prison Experiment simulation and the Hamariyah Iraqi Village simulation. The events in these data sets were the outcomes of the actions that agents took within each respective simulation. The Hamariyah simulation only allowed one event at a time, so attention in this system also follows Equation 7.5. The Stanford Prison Experiment involves multiple simultaneous events, so salience relates to attention through the more general form noted in Equation 7.4. However, even in this case, each input to attentional salience has a positive contribution to salience and should contribute positively and their relative importance on attention should be maintained.

As such, the regression provides a useful system verification measure. If the regression weights from in the data ($\overrightarrow{\beta}$) match weights set in the system ($\overrightarrow{w}$), the regression demonstrates that each input has the appropriate sign and importance for determining attention. Since the relative importance of weights was of interest, the regression intercepts were calculated but were not reported as they have no value for interpreting the results. Multivariate linear regressions were performed using the "bigglm" R statistical package, intended for large generalized linear models (Lumley, 2009).

7.2.3 Mann-Kendall Trend Tests

Mann-Kendall trend tests were used to determine if certain time series tended to be negative or positive over time. For example, the Stanford empirical data posits a number of emotional trends, such as that prisoners' emotions (as a group) became negative over the course of the experiment. The Mann-Kendall test is a non-parametric trend test, which analyzes a time series of values. The null hypothesis for the trend test is that the data series consists of independent and randomly ordered values. This test uses the Mann-Kendall statistic to calculate the significance of a time-dependent trend and the direction of the trend. Mann-Kendall tests were performed using the R statistical package "kendall" which implements the Mann-Kendall trend test.

In a respect, a standard Mann-Kendall trend test is non-ideal for emotions. Emotions typically involve some level of cyclic behavior, which will reduce the significance level of the basic Mann-Kendall trend test. However, the ground truth statements from the Stanford Prison Experiment papers do not describe emotional cycling of moods- they describe basic trends. For this reason, the basic Mann-Kendall test was used. With that said, the original experiment's mood trends were based upon only two or three data points for each subject so the confidence of the ground-truth data makes this analysis harder to interpret. Despite these limitations, examining the correspondence of the Stanford Prison Experiment simulation trends compared to the real experimental trends was an interesting avenue to examine as part of the external validity testing.

7.2.4 Meme First Expression Ordering

First expression ordering analysis is a metric selected for externally validating meme transmission against external data. The first time an agent expresses a meme provides direct proof that an agent has learned a meme. While the observed data may not provide the exact time the agent learned the meme, their first expression provides an upper bound for the time that each agent learned the meme. By definition, an agent A must know the meme at some t_A value before the time of an agent's first expression T_A, where $T_A > t_A$. As such, the order of first expressions provides a metric that bounds the time span that each agent could have learned the meme. This means that each agent can be ordered by their time of first expression, as shown in Equation 7.6 (where it is assumed agents are assigned a subscript according to their order of expression).

$$T_{A_1} <= T_{A_2} ... <= T_{A_N} \tag{7.6}$$

If the cumulative probability of expression is an increasing function with respect to the time passed after learning a meme, then the first expression time provides information about the order that the each agent learned the meme. In some circumstances, this information may be harnessed to estimate the order that agents learned memes. This approach will not be used in this analysis, however, since individual factors have a considerable impact upon the time between learning a meme and expressing a meme (if it is expressed at all).

Instead, the times of first expression are considered as an ordering that ranks a covariate combination of an agent's learning of the meme and their motivation to express the meme. As noted in Section 6.1.2, an order of first expression for three potential memes was extracted from the Stanford Prison Experiment hourly logs. Similarly, each simulation run produces an ordering for when agents first performed each action. Comparing the simulated orderings against the observed orderings provides a form of external validation.

Inversion Count Algorithm

In principle, it is straightforward to compare two ordered series against each other statistically. Any ordered series can be reduced to a set of ordered pairs, and the sets of ordered pairs can be compared directly, with adjustments made to adjust for duplicate information. In practice, the situation for meme expression is much messier. Agents may not express a meme on a particular run, leading to ambiguities at the tail of the ordered series. There is simply no way to infer which agent expressed a meme first if neither expressed it. Additionally, even if the series were always complete- it would be necessary to adjust for duplicated information due to transitivity (ex. A > B and B > C implies A > C).

As a result, an algorithm was developed to statistically analyze the distance between two ordered series, which allows for right censoring of both series and for ties in both series. This algorithm is based upon the principle of series inversions. An inversion count algorithm can determine the minimum number of single-element swaps that are necessary to turn one ordering into another ordering. Table 7.5 displays a simple example of inversion counting. Such algorithms are frequently used to measure the distance between sequences, such as in DNA chains. The inversion number of a random permutation follows a distribution somewhat similar to a normal distribution (Margolius, 2001). The mean inversion count for random permutations is half the maximum inversion count, giving a null-hypothesis condition when examining inversion counts.

Table 7.5: Inversion Counting Example

	Sequence	Inversion Tabulation
Real Sequence	[A,B,C]	-
Permutation	[C,B,A]	-
	[B,C,A]	+1
	[B,A,C]	+1
	[A,B,C]	+1
Inversion Count		**3**

The algorithm takes advantage of these principles by calculating the inversion number to turn a simulation sequence into the ground-truth sequence and comparing it to the maximum possible number of inversions possible, given the simulation sequence and ground truth sequence. The algorithm handles ambiguously ordered or simultaneously occurring events by removing ignoring inversions within that subsequence when calculating the inversion number and the maximum inversions. This retains the property that an above average number of inversions would be more than half of the maximum possible inversions.

Table 7.6 shows the results of using the modified inversion distance algorithm on some example sequences. For the sequence and permutation, the second

Table 7.6: Modified Inversion Count Examples

Sequence	Permutation	Inversions (I)	Max Inversions (M)	Nearness (1 - I/M)
[A,B,C,D]	[A,B,C]	0	3	1.00
[A,B,C]	[B,C] (A)	2	3	0.33
[A,B,C]	[B] (A,C)	1	2	0.50
[A,B] (C,D)	[C,B,D] (A)	4	5	0.20

parenthetical list represents right-censored elements (with an unknown order). These examples demonstrate some of the dynamics of the distance calculation. The first example has no inversions, as the sequence is in the correct order. While one element is not present, it is not considered unless one designates it as being censored in some way. The second example demonstrates what happens when an element is censored from the permutation sequence. Since only one element is censored, the sequence might as well be fully observed (since the order is fully known). The third example has two right censored elements, meaning there is one less observable inversion. In this case, both the number of inversions and the number of possible inversions are reduced by one. This reduces the distance and improves the nearness score, since the inversion between A and C in the prior example has been replaced by uncertainty. The last example demonstrates the ability to have censored elements in the ground truth sequence. Ties are handled similarly to censored elements, with no inversions counted by either the inversion count nor the maximal inversion count. Appendix I notes some additional properties of the algorithm. As noted, given a random permutation with random right-censoring (the null hypothesis), the nearness calculation approaches 0.5 for this metric. A nearness above 0.5 means that a sequence is closer than chance.

This modified inversion number algorithm provides a useful metric for comparing the distance between an individual simulation sequence against the ground truth, while naturally normalizing this distance and adjusting for censored data and ties. It provides a way to determine which experimental cases are closer to the ground truth data, and a way to tell if the simulation as a whole is performing better than chance at predicting the order of first expression for each meme.

Median Expression Position

A second metric for comparing simulation expression orderings with ground truth orderings is by determining the median order for each agent's first expression. This ordering is determined by calculating an agent's order within a given simulation run, as compared to its peers. For each agent, this generates a set of data in the form $[O_1, O_2, ..]$ where O_1 is the order that the agent took the action in the first run and O_2 is its order for the second run, etc. Due to the possibility of ties, an agent may share its order with another agent on a given run. In this case, all simultaneously acting agents are assigned to the average of the slots they would have occupied as a group (ex. three agents tied for fourth would all be marked as 5, the mean of [4,5,6]). From an agent's expression position across multiple simulation runs, an agent's median expression position can be calculated by taking the median value.

Table 7.7 displays a set of 3 example orderings and their resulting median sequence. For reference, the indices of the sequences are shown in the first column. As one can see, the positions in the median sequence are determined by the median position of each term. A is the first element, since its positions were (1,1,4). C and D share the third position, since their positions were (3,3,4) and (2,3,4) respectively. This approach helps generate a typical ordering for the elements, which is representative of those observed in the individual sequences.

Table 7.7: Median Sequence Example

Index	Sequence 1	Sequence 2	Sequence 3	Median Sequence
1	A	B	A	A (1)
2	B	D	B	B (2)
3	C	C	D	C, D (3)
4	D	A	C	

Using the median values of agent's expression positions, an expression order can be generated that indicates the typical positions that agents first expressed a meme. This provides an alternative method for comparing the simulation orderings against the holdout data. It also provides insight into which agents typically did not express a meme, since their median expression position will be "Never." Additionally, since this produces an expression ordering the inversion count method can be applied to the median position ordering as well.

7.2.5 Diffusion Rate Analysis

Measuring the spread of affordances can be treated as a diffusion of innovation problem. The Innovation Decision Process (IDP) theory can be used to frame this analysis. This theory has five stages: knowledge (awareness), persuasion, decision, implementation, and confirmation (Rogers, 1995). These parameters are present within the simulation framework, as shown in Table 7.8. This means that the diffusion analysis focuses on the Knowledge and Implementation phases of the IDP theory, noted by asterisks in Table 7.8. This analysis considers three possible states for PMFServ agents with respect to a meme: unaware, aware but not expressing (knowledge), and expressing (implementation). Learning occurs when agents move from unaware to aware. Expression occurs only when an agent actively engages in the afforded action. Measuring the reproduction dynamics of memes requires monitoring agents as they transition between these states.

Table 7.8: Innovation Diffusion Analysis Metrics

IDP Theory Stage	**Measurement**
Knowledge*	Existence of meme in agent memory
Persuasion	Utility of expression at each time step
Decision	Decision choice of agent
Implementation*	Action implementation and result
Confirmation	Transitions between expression and non-expression

Classical diffusion of innovation models have only two states, adopters and non-adopters (Rogers, 2002). This is based upon the assumption that an adopter not only adopts a practice but continues to use it. Since memes are only intermittently expressed, this is not a good assumption for simulation. Instead, diffusion is examined separately at the learning level and the expression level. For each of these properties, the total count of agents who have reached these states is plotted over time to examine the diffusion curve and adoption rate. This provides insight into the rate of acquisition and of initial adoption. This allows comparing the simulation diffusion curves against the classical S-curve of adoption from Rogers (1962).

7.2.6 Granger Causality Test

Related to diffusion, the time-causality between learning and first expression is of interest. Clearly, learning is necessary for expression. Likewise, expression is necessary for new learning- a chicken and egg problem. While learning and expression must be mutually causal, a first expression is not necessary for learning because later expressions can also promote learning. With that said, the first expression by an agent may be more causal for learning due to its potential to reach new agents who are socially well-connected to the agent who is expressing for the first time. It is worthwhile to analyze if learning and first expression are significantly causal to each other within these simulations, as well as to examine the immediacy of this causality (i.e. are people expressing right after learning, or much later).

The Granger Causality test can perform this sort of analysis because it is a statistical test that determines the likelihood that one time series ($\overrightarrow{X}$) is causal to another time series ($\overrightarrow{Y}$), based upon a fixed lag factor (Granger, 1969). A Granger Causality test is a special type of regression. First, an autoregression is performed on the time series of the caused variable ($\overrightarrow{Y}$) to capture the variance of prior terms in the series in predicting the current value, where prior terms up to the lag factor are considered. Next, a second regression is run using the significant prior terms of $\overrightarrow{Y}$ and also the corresponding prior terms from $\overrightarrow{X}$. A statistical test is applied to the final regression where the null hypothesis is that none of the lagged terms from $\overrightarrow{X}$ add to the explanatory power of the regression. Effectively, this test checks if the prior terms from the time series $\overrightarrow{X}$ help predict $\overrightarrow{Y}$ as opposed to terms from $\overrightarrow{Y}$ alone.

The Granger Causality test was applied to test the causality of learning on first expression and on the causality of first expressions on learning, for each simulation run in the Stanford Prison Experiment. For each causality test, the lag factor was varied over a fixed range of lags representing time values from 10 minutes of simulation time later to one hour of simulation time later.

If learning appeared strongly causal to first expression, this meant that agents tended to express a meme not long after learning it. If first expression was causal to learning, that indicates that new agents expressing a meme are reaching otherwise resistant agents- ones who had not learned from other agents expressing the meme. This second effect is of particular interest- the possibility that an agent's first expression helps reach previously unreachable agents. This sort of effect is observed in real life. For example, if a friend with no taste in movies constantly recommends a movie they may be ignored repeatedly. However, if they get someone whose opinion you trust to see the movie and the trusted person recommends it, you may pay attention to the title of the movie.

To explore these causality relationships, the Granger Causality test was chosen (Granger, 1969). The particular implementation used was the "grangertest" implemented in the R "lmtest." This variant uses a Wald model-comparison test in the background, which allows either F-test or Chi-Squared test statistics to be employed. For these analyses, the F-test variant was used. From these tests, the probability of a causal relationship in each direction was examined.

7.2.7 Sub Group Analysis

Of key interest are the differences between agents who tend to learn certain memes and express certain memes. To examine these differences, each agent's time of first learning and time of first expression were calculated across all of its simulation runs during a specific experimental condition, for each meme. The first learning was the first time that the agent acquired the meme. The first expression time was the first occurrence when an agent performed the meme action. These two parameters were used to segment the total agent population into subgroups based on their learning and expression rates.

Two methods were used to segment agents into subgroups. The simplest approach examined each property separately and segmented agents into quartiles.

Quartiles are a common approach used for simple classification based upon a single property. The highest and lowest quartile groups were compared against each other. So then, attempts were made to discover the differences between agents who most frequently expressed a certain meme as compared to those who seldom expressed the same meme. This method was applied to the Stanford scenario, where the number of agents was small.

The second method of segmenting agents into groups was a model-based clustering algorithm which employs normal mixture modeling. Clustering was applied to the Hamariyah Iraqi village because the number of agents was significantly higher, making it infeasible to break the agents down into quartiles and examine each agent individually. Additionally, clustering techniques have the benefit of being able to handle more than one variable simultaneously. This allowed categorizing agents with respect to both their speed of learning and first expression.

Normal mixture model clustering was chosen over a variety of other clustering techniques available. Normal mixture clustering tends to generate elliptically shaped clusters for two dimensional data. From looking at the raw data plots, many of the Hamariyah experiments appeared to have clusters of that approximate shape. As such, normal mixture model clustering was chosen to classify agents into subgroups. K-means clustering was considered as an alternative but was not used due to its strong dependence on the number of clusters selected. Given the solid effectiveness of clustering using normal mixtures, k-means were not needed to classify agents in the Hamariyah Iraqi village.

The clustering algorithm used for clustering was mClust, a package implemented in R (Fraley & Raftery, 2003). mClust was chosen because it is a flexible and tested normal mixture model clustering package. mClust was used to generate clusters among the agents based upon their average learning time and average first expression time for memes. The mClust algorithm takes a matrix where each row i represents a data point in the form (x_i, y_i, ...). For this analysis, each data point represents different properties of the same agent. This means that each row of the matrix is in the form (FirstLearningTime of $Agent_i$, FirstExpressionTime of $Agent_i$). Based upon this matrix, the mClust algorithm uses Bayesian techniques to apply the appropriate mixture model, assign points to clusters, calculate the cluster means, and calculate a variety of other factors such as uncertainty of assignments. This software also provided built-in graphing capabilities which were used to generate the cluster maps presented. A cluster analysis was completed for each meme (GiveInformation and PlantIED) under each experimental condition (Hypothesis and Random).

Exploratory analysis of the generated clusters was performed to discover reliable differences between clusters, based upon agent properties and contexts.

Cross cluster analysis was performed to compare the distributions of agent property values between clusters. The properties examined were the group memberships of the agents, the agents' opinions of other groups, their employment levels, their authority, and their personality traits (from the GSP Tree, as explained in Section 5.2). ANOVA analysis helped detect properties that had significant differences between other clusters. A Scheffe post-hoc test was run to determine the significance of differences between each pair of clusters, for each property. These differences are discussed to explain why certain clusters were early or late adopters, for either learning or expression. The ANOVA analysis was completed using the SAS statistical software.

7.3 Internal Validity

Internal validity testing is an important part of verifying that the agent cognitive model works as expected. This testing is intended to make sure that the computational model works as described in Chapter 5. The agent cognitive model is constructed from a set of interacting cognitive components, as explained in Chapter 5. Two types of internal validity testing were performed, as noted in Table 7.9.

Table 7.9: Internal Validity Analyses Performed

Analysis	**Findings**
Test the relationships in cognitive model	Components that affect attention and learning work in the correct direction (i.e. match the social science studies used to extend the PMFServ agent cognitive model).
	Components that affect attention and learning work have the correct relative weights (i.e. match the attention salience weights assumed in Table 5.7).
Check if the same relationships are observable in the simulated scenario data	Cognitive component relationships are unclear using same analyses, due to collinearity between cognitive components.
	Situational factors significantly affect which components covary.
	Novelty's relationship with attention is masked due to negative feedback between novelty and learning.
	Ingroup membership and valence tended to covary (people tend to like those in their ingroups).

The first test is a simple sanity check, that proves the models are implemented correctly. This first check tests that the computer code to implement the model is working properly. The second part of internal validity analysis examines how the components appear to interact, from data collected from the Stanford Prison and Hamariyah simulation data. This second analysis is intended to test if the implemented relationships would be obvious within the collected data from an experiment. This test examines if the regression approach would be appropriate for examining these relationships, if human attention incorporated these factors in a similar way. This is important because the current linear weights of the inputs to attentional salience are "best guess" weights. If an experiment could be run to infer these weights experimentally, the cognitive model could be substantially improved. As such, the simulations provide a useful testbed for examining if this sort of regression might help determine better weights for the attention salience.

Internal validity testing differs from external validity testing in two ways. Firstly, internal validity testing only checks that the PMFServ agent cognitive model works as intended. They are compared against the assumptions and empirical relationships used to build the model. On the converse, external validity testing compares simulation outcomes against relationships not directly used to construct the cognitive model or scenario.

This test ensured that the cognitive model for agents oriented attention consistently with the weights given to each component of attentional salience, as defined in Section 5.2.5. This internal validity analysis used a single agent with a special test framework, rather than a full scenario. The importance of each salience factor for attention was calculated using a logistic regression and compared against the weights defined in Table 5.7, which shows the weights given for each factor into the attention salience calculation. This ensures that the agent's cognitive model properly implemented the empirical relationships used to design cognitive components. This analysis found that the agent cognitive model did handle learning and attention as intended.

Secondly, the same regression analysis was performed on the collected affordance transmission data collected from the simulations, as defined in Table 7.4. Since the same underlying cognitive models are used by the simulations and the test framework, the underlying weights are also the same. However, performing this analysis showed that in a more complex scenario it is much more complicated to determine the relative weights. This is because in both scenarios, certain structural elements created collinearity between the different components of attentional salience. For example, even though cognitive model considered valence and ingroup membership to be independent, both scenarios showed a strong correlation between these factors. Certain interesting findings were found in this analysis, which are noted. These findings indicate that even if attention was handled similarly to the proposed model, inferring the values of these weights would be non-trivial.

7.3.1 Verifying Event Salience Component Weights

As noted in Section 5.2, attention and learning are mediated by a set of cognitive models that model social influences (authority, conformity, ingroups, reference groups, similarity, valence), action characteristics (motivated attention, transferability), and general attentional factors (novelty, selection) that affect the likelihood that an agent will learn from an observed event. Each of these factors was chosen because it was based upon an empirical study which showed that the factor had a positive correlation with the probability of being able to recall a message or event. This analysis verifies that the components of the cognitive model capture the relationships they are intended to model.

Table 7.10: Event Salience Component Weights (Copy of Table 5.7)

Component	Assumed Weight	Source	Process
Authority	0.33	Mantell (1971)	Peripheral
Conformity	0.34	Tanford and Penrod (1984)	Peripheral
In-Group	0.30	Tajfel (1982)	Peripheral
Motivation (central)	0.47	Roskos-Ewoldsen and Fazio (1992)	Central
Novelty	0.21	Johnston et al. (1990)	Mixed
Reference Group	0.30	Kameda et al. (1997)	Peripheral
Selective Attention	0.32	Simons and Chabris (1999)	Mixed
Similarity	0.47	Platow et al. (2005)	Peripheral
Transferability	0.10	Bandura (1986)	Central
Valence/Halo	0.38	Hilmert et al. (2006)	Peripheral

Earlier in Chapter 5, the cognitive components were described which are used to determine if an agent learns from an event that they can physically observe. Learning from an event requires that an agent must pay attention to the event, after which learning occurs probabilistically. As noted, under the current settings all attended events result in learning. As such, attention is the key factor that controls which events are learned. Attention to an event is determined probabilistically as a function of the attentional salience of that event. Table 7.10 displays the assumed importance of each cognitive component in determining if an event is attended. Attentional salience for an event is a weighted sum of each of those inputs, as weighted by the displayed weight.

As mentioned earlier, these weights are not necessarily empirically true but are modeling assumptions that form the “best guess” from examining the stated studies. As noted in Section 5.2.5, no empirical study has examined all these factors simultaneously. As such, the linear structure of the attentional salience calculation and the weights must be considered as modeling assumptions. The true structure and relative importances of these factors in determining attention and learning still has considerable ambiguity. This is an area where more empirical research could significantly improve the model quality.

However, the calculation of attentional salience does capture some important information. This analysis is intended to test for the following:

1. Firstly, according to the empirical studies, all of the factors in Table 7.10 should have a positive relationship with learning from an event.

2. Secondly, the computational agent cognitive model components should affect attention with these relative weights. This is an important sanity check that the cognitive model is wired correctly.

3. Finally, this analysis should demonstrate that relative weights can be calculated based upon observable data without needing to know their values a priori. This is important because it shows that an empirical study could be designed that would improve these weights, if attentional salience could be approximated by a sum of linearly weighted components.

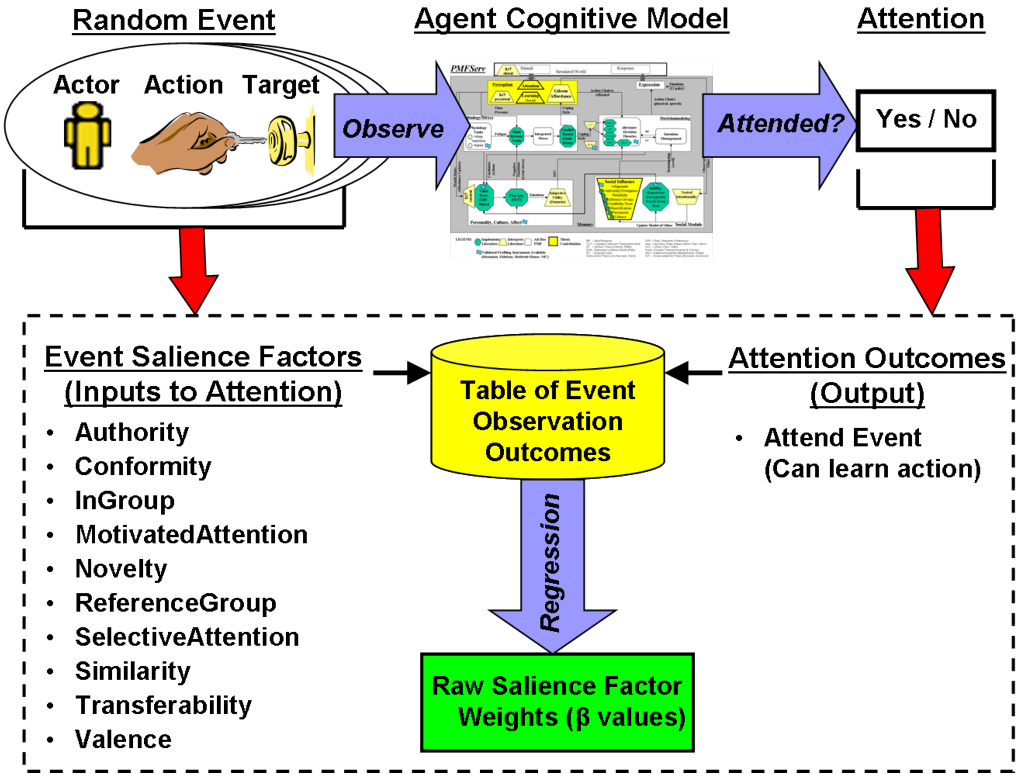
Separate from the simulation runs, a test was made using the attention model alone. Events were passed to an agent for which each salience component was selected from a uniformly random distribution in [0,1]. A set of one hundred thousand randomly generated events were passed to the attention model, with the outcome recorded (attended vs. not attended). This attention model used an inattention salience of 8.0 (same as Hamariyah) and presented one event at a time.

A test system was implemented which examined an individual agent's attention by passing it events with randomized salience component inputs. Figure 7.1 displays the system that was used to test that the agent's cognitive model was handling attention correctly. The attention model was shown a series of individual events, each of which was specifically designed to have a particular value for Novelty, Similarity, and all other components used for calculating attention salience as noted in Table 7.10. A data set was generated by presenting one hundred thousand events to the attention model, where each event had random and independent inputs to the salience calculation.

$$\begin{aligned}
Attended(e) = &\beta_0 \cdot \text{Novelty(e)} + \beta_1 \cdot \text{MotivatedAttention(e)} + \\
&\beta_2 \cdot \text{SelectiveAttention(e)} + \beta_3 \cdot \text{Transferability(e)} + \\
&\beta_4 \cdot \text{Authority(e)} + \beta_5 \cdot \text{Conformity(e)} + \beta_6 \cdot \text{Similarity(e)} + \\
&\beta_7 \cdot \text{Valence(e)} + \beta_8 \cdot \text{InGroup(e)} + \beta_9 \cdot \text{ReferenceGroup(e)} + m + \epsilon
\end{aligned} \tag{7.7}$$

The data set was processed using a logistic regression, shown for reference in Equation 7.7 (a copy of Equation 7.3). This regression returns raw regression

Figure 7.1: Event Salience Test Setup



β weights. These raw β weights do not directly correspond to the true salience coefficient weights shown in Table 7.10 because attention is the output variable, rather than attentional salience. As noted in Section 7.2.2, attention is a probabilistic function of the sum of salience terms for observed events as well as inattention salience. While the raw β weights will not match the true salience weights, if each set of weights is normalized to sum to 1, then they should match up.

For example, the maximum attention salience occurs when all inputs are equal to 1 and sums to 3.22 (the sum of the weights in Table 7.10), with the contribution from novelty being 0.21. If the raw β weights sum to 6.44 (excluding the intercept), then the β weight from novelty should be 0.42. If the cognitive model is working properly, then Equation 7.8 should hold for the salience input weights where β_i represents a raw β regression weight and w_i represents the true salience input weight for the input i. If this holds, the test has demonstrated that the agent pays attention to events because of the factors in Table 7.10 and that the relative importance of each factor corresponds to its salience weight.

$$\frac{\beta_i}{\sum_i \beta_i} \approx \frac{w_i}{\sum_i w_i} \qquad \forall i \in \{SalienceInputs\} \tag{7.8}$$

To make this easier to examine, the raw regression weights for attention are multiplied by a factor of $\frac{\sum_{w_i} |w_i|}{\sum_{\beta_i} |\beta_i|}$. While not changing the importance of the weights relative to each other, it allows them to be directly compared against the underlying weights in the attention salience model (True Coefficients). The results of the regression are presented in Table 7.11. This table shows the raw β coefficients, the rescaled β coefficients, the true model coefficients, and the difference between the rescaled β weights and the true salience weights.

Table 7.11: Component Weights for Attention (Random, Indep. Components)

Salience Input	**Raw Coefficients (β_i)**	**Rescaled Coefficients**	**True Model Coefficients (w_i)**	**Rescaled - Actual**
Authority	0.21	0.34	0.33	0.01
Conformity	0.22	0.36	0.34	0.02
InGroup	0.18	0.29	0.30	-0.01
Motivation	0.27	0.44	0.47	-0.03
Novelty	0.13	0.22	0.21	0.01
Reference Group	0.19	0.31	0.30	0.01
Selection	0.20	0.31	0.31	0.00
Similarity	0.30	0.48	0.47	0.01
Transferability	0.07	0.11	0.10	0.01
Valence	0.23	0.37	0.38	-0.01

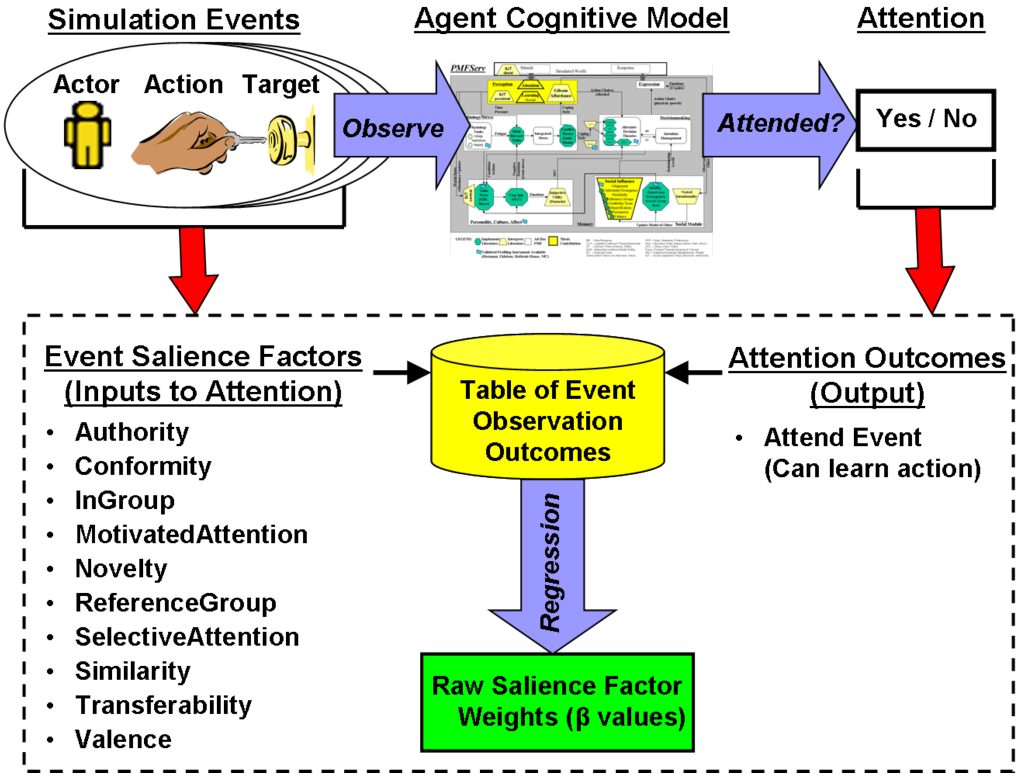
This verifies that the attention model integrates the attention salience components as designed. Each component has a measurably positive contribution toward learning, meaning that the cognitive components are correctly implemented. It also shows that, given the model design, a logistic regression can tease out the relative impact of different inputs to the attentional salience despite its low predictive utility. This means that if human attention was directed by such a process, the relative weights could be inferred.

7.3.2 Event Salience Component Weights (Simulation)

A logistic regression analysis was also done on the data from a case from the Hamariyah scenario (Randomized Condition) and from the Stanford scenario (Hypothesis Condition). The analysis performed was similar, with the only difference being the source of the events analyzed. While the prior case examined specially generated events which randomized all inputs to attentional salience, the events for these conditions resulted from the actual running of the simulation.

Likewise, whether the attention paid attention to that event was the real outcome within the simulation during that run. Figure 7.2 displays the setup used to examine the salience component weights from the simulation data.

Figure 7.2: Event Salience Simulation Data Collection and Regression



The Hamariyah Iraqi village operates similarly to the randomized test condition. Agents observe only one event at a time, which competes with the same level of unrepresented noise (inattention salience of 8.0). The Stanford condition is different due to the fact that agents process multiple events simultaneously and can perceive up to 4 events. The inattention salience is small, so multiple events will be typically be perceived. However, this will result in some interaction between events, since events are competing for attention against each other.

Table 7.12 displays logistic regression weights estimated from the Hamariyah village scenario and the Stanford scenario. The Hamariyah regression is calculated over approximately 5 million observations where an agent could perceive another agent taking an action, while the Stanford regression was taken over approximately 2.75 million observations. This large number of observed cases for each is a result of the number of observations being a function of: # runs × # actions per agent × $(\#agents)^2$. Only one condition for each scenario is presented, as they are representative of the other cases using the same scenario.

Table 7.12: Actual and Regression β Weights for Attention

Salience Input	**Actual Model Coefficients**	**Stanford (Hypothesis) Coefficients**	**Hamariyah (Randomized) Coefficients**
Authority	0.33	0.08	0.24
Conformity	0.34	0.61	0.72
InGroup	0.30	0.42	0.13
Motivation	0.47	-0.24	-4.31
Novelty	0.21	0.06	-0.16
Reference Group	0.30	-0.04	0.47
Selection	0.31	0.26	0.36
Similarity	0.47	0.13	0.41
Transferability	0.10	0.82	-0.34
Valence	0.38	0.80	0.51

It is evident that these weights in a real simulation differ significantly from the fully random condition. Since the cognitive model hasn't changed at all between the Hamariyah condition and the random test, the difference comes from the inputs: there is no longer an assurance that the inputs are independent. This indicates that the issue is multicollinearity- the inputs covary and are explaining the same variance in attention. To test if collinearity between factors made it impossible to infer the weights using a regression, two analyses were applied. The first analysis was a set of regressions applied on subsamples of the simulation data. This helped identify which factors were unstable or had other sampling bias that confounded the full regression. The second analysis was a correlation analysis to examine covariance between each pair of factors. This analysis helped explore which factors might have interacted.

7.3.3 Subsample Regression Analysis

The subsample regression analysis was performed to check for collinearity or time-dependence in the observations. If collinearities exist, these should result in large swings in the β weights for the inputs that are collinear. Since the simulation data set size is large (millions of observations), it is possible to run regressions on multiple random subsamples of observations to check for such instability. To check for time-dependence, a series of regressions was performed on evenly-sized time intervals of the full simulation data sets. The data for each time slice was drawn from the same time range across many runs. By subsampling over time, this analysis examined if the β weights appeared to have a trend through the simulation.

Detecting Collinearities: Random Subsample Regression Analysis

The random subsample analysis completed utilized the same form of regression displayed previously in Equation 7.7. It also used the same data sets used earlier in Section 7.3.2, the Hamariyah Randomized Condition attention data and the Stanford Hypothesis Condition attention data. However, instead of computing a single regression each full data set, this analysis computed a number of regressions on subsamples of observations from each full data set. This was intended to expose unstable β terms in the regression that might result from collinearity.

Each smaller regression consisted of 10,000 observations. These observations were sampled from the full data set under that condition, which consists of observations from numerous different runs and observations within those runs. The subsamples were generated independently, sampled from the full data set under that condition. As such, each of these random subsets is a limited but representative sample of the full set of observations ($\overrightarrow{O}$) used to calculate the regression coefficients shown in Table 7.12.

$$\overrightarrow{O}_{Sample(k)} = choose(\overrightarrow{O}, k) \tag{7.9}$$

$$P(\overrightarrow{O}_{Sample(k)} = x) = \frac{1}{\binom{N}{k}} \tag{7.10}$$

Each subsample was obtained by using a Choose k algorithm, where each subsample consists of k observations sampled from the full set. In generating a single subsample, the observations are sampled without replacement (a subsample will not have any duplicates). Across samples, there is no restriction on re-using elements, all subsamples are generated starting with the full set of observations. So then, assume each subsample ($\overrightarrow{O}_{Sample(k)}$) consists of k elements from the full set ($\overrightarrow{O}$) that consists of N elements. The function to select a subset can be stated as in Equation 7.9, where the probability of any given subsample is noted in Equation 7.10.

Since these subsamples were random and independent, it is possible that the same observation might be included in one or more subsamples. However, such cases are unlikely due to the size of the subsample sets (10,000 observations) compared to the full data sets (2.5 to 5 million observations). Since these less powerful regressions should still have enough power to be representative, high instability in the β values imply collinearity issues.

Regressions were applied to 100 subsamples, where each sample consisted of 10,000 observations. Table 7.13 displays the mean and variance of the regression β values for each of the attention salience inputs, for subsamples of the Stanford Hypothesis and Hamariyah Randomized conditions (same conditions used for the full simulation regressions). Looking at these regressions, it is evident that significant instability exists for all coefficients and major instability exists for

Table 7.13: Subsample Regression β Weights for Attention

	Stanford Hypothesis		**Hamariyah Randomized**	
Salience Input	**Average β**	**StDev β**	**Average β**	**StDev β**
Authority	0.26	0.13	0.20	0.11
Conformity	0.09	0.13	0.88	0.06
InGroup	0.41	0.14	0.08	0.09
Motivation	-0.23	0.27	-3.68	0.28
Novelty	-0.01	0.29	-0.30	0.11
Reference Group	-0.02	0.74	0.54	0.31
Selection	0.83	0.13	0.42	0.42
Similarity	0.63	0.26	0.42	0.08
Transferability	0.15	0.13	-0.43	0.05
Valence	0.79	0.16	0.56	0.08

some coefficients. From looking at these trends, the majority of β weights have standard deviations greater than 0.1 in both conditions. This indicates that collinearity is a problem for the regression, leading to unstable β weights. Additionally, the regression algorithm itself seems to tend toward assigning more extreme β weights rather than distributing weights more evenly. This is a documented issue with typical regression approaches in the presence of multicollinearity (Lipovetsky & Conklin, 2001). As such, even the seemingly stable weights may be improperly biased. Given that collinearity appears to be an issue, the relationships between the factors will be examined in Section 7.3.4.

Detecting Time Trends: Time-Interval Regression Analysis

To detect issues resulting from time-dependent interactions, the full data sets were also split into subsamples based on their simulation time step values. Unlike the prior analysis, these subsamples are deterministic and non-overlapping- no observation from one subsample could be present in another subsample. This analysis also used the same regression displayed previously in Equation 7.7, as well as the data sets used earlier in Section 7.3.2 (Hamariyah Randomized Condition and Stanford Hypothesis Condition). A Mann-Kendall trend test was applied to the β coefficients for each salience input, examining the trend of each coefficient across different periods of each simulation.

The subset samples for the time interval regression were created by splitting up the full set of observations based upon their simulation time step value (listed as “Step” in Table 7.4). This value determines when the action occurred during the simulation. For each time step, multiple observations exist. This is because there are multiple runs, meaning the same time point occurs in each simulation. Additionally, many agents can observe the same action and determine if they pay attention to it. This allows even a small slice of simulation time to have many associated observations.

The observations for each simulation condition was split into subsets that covered time intervals that approximately equal in time length. Each of these subsets consisted of data from multiple simulation runs that started using the same experimental condition. Assume $\overrightarrow{O}_r[t]$ represents the observations for run r that occurred at the Step value (t). If the time interval length is assumed to be L and there were Z runs in that condition, the data will be split into subsamples as shown in Table 7.14.

Table 7.14: Time-Interval Subsample Approach

	Step (t)	Run 1	Run 2	...	Run Z
Subsample 1	1	$\overrightarrow{O}_1[1]$	$\overrightarrow{O}_2[1]$	...	$\overrightarrow{O}_Z[1]$
	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
	L	$\overrightarrow{O}_1[L]$	$\overrightarrow{O}_2[L]$	...	$\overrightarrow{O}_Z[L]$
Subsample 2	L+1	$\overrightarrow{O}_1[L+1]$	$\overrightarrow{O}_2[L+1]$	...	$\overrightarrow{O}_Z[L+1]$
	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
	2L	$\overrightarrow{O}_1[2L]$	$\overrightarrow{O}_2[2L]$	...	$\overrightarrow{O}_Z[2L]$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	...	$\vdots$

A logistic regression, in the same form as shown in Equation 7.3, was applied to each of these subsets to estimate the β coefficients for each input to salience. For each input to salience, the associated β coefficients were considered as a time series based upon the time period for the subset. For example, the Hamariyah regressions assigned Ingroups a β value of 0.32 the step 1 to L subset, a β of 0.26 for step L+1 to 2L subset, and so on. If the distribution of observations is time-invariant, there should be no significant trends in the β coefficients for each of the inputs to attention. To test for this, a Mann-Kendall trend test was applied to each β time series to determine the association (trend direction) and the probability of the null hypothesis (no trend).

For the Hamariyah Randomized data set, the data was split into slices 72 steps in length ($L = 72$). This length was chosen because it was the number of steps required for each agent to take one action. This created 48 non-overlapping subsets, each with approximately 100,000 observations. Table 7.15 displays the Mann-Kendall trend analysis of the regression β values calculated for each time interval. In this table the Tau coefficient displays the direction and strength of the trend, while the p value indicates the probability of the null hypothesis (no trend). From looking at this table, it appears there are potentially significant trends occurring over time. Motivation, novelty, and valence appear to experience a negative trend in their β values over time, making them less indicative of attention. All the other factors appear to have increasing coefficients, except for InGroups, which appear to have a stable influence. This indicates that the collinearities between inputs experience some time-dependent trends. The next section, which explores correlations between factors, examines why some of these trends might occur.

Table 7.15: Interval Regression β Weights for Attention (Hamariyah Randomized)

Salience Input	**Tau Coefficient**	**p**
Authority	0.17	0.09
Conformity	0.19	0.06
InGroup	0.10	0.34
Motivation	-0.56	$2 \cdot 10^{-8}$
Novelty	-0.17	0.10
Reference Group	0.17	0.08
Selection	0.27	0.01
Similarity	0.17	0.10
Transferability	0.20	0.04
Valence	-0.28	0.01

The Stanford Hypothesis condition data was also analyzed using a similar technique. Each time slice consisted of 12 steps, creating 58 subsets which each contained approximately 50,000 samples. However, due to the changing guard shifts, it is harder to interpret the β coefficients. Since a change in guard shifts may change some of the collinearities, there will be a degree of periodicity in the β coefficients. In order to accommodate this, a Seasonal-Kendall trend test was applied. This test works similarly to the Mann-Kendall test, except that it preprocesses each season by finding the median and then looks for trends across the seasons. For this analysis the guard shifts were treated as “seasons” of the day, so that a 24 hour day consisted of 3 seasons (guard shifts). While this technique loses some temporal information, it is necessary to accommodate the periodic nature of guard shifts. The trend test implementation used was developed by the US Geological Survey (USGS) and its details are described in Helsel, Mueller, Slack, and Geological Survey (US) (2006).

Table 7.16 displays the trends in the β coefficients for the Stanford Hypothesis condition attention components. The Stanford Prison experiment appeared to have less clear time-dependent trends, possibly due to the periodicity of the guard shifts. However, it still showed a few strong trends in the β values over time. Valence and reference group influence appeared to have less power for attention over time, while similarity appeared to increase its β values significantly. Overall,

Table 7.16: Interval Regression β Weights for Attention (Stanford Hypothesis)

Salience Input	Tau Coefficient	p
Authority	-0.13	0.67
Conformity	-0.07	0.89
InGroup	0.07	0.89
Motivation	-0.20	0.48
Novelty	-0.20	0.48
Reference Group	-0.53	0.03
Selection	0.40	0.12
Similarity	0.87	0.01
Transferability	0.27	0.32
Valence	-0.80	0.01

this data is harder to draw clear conclusions from but it indicates that some time-dependent changes in collinearity occurred in this simulation as well.

Looking at the Tau correlation coefficients from both experiments, a few trends seem to emerge. Firstly, the InGroup β values appear to be fairly consistent with respect to time- neither increasing or decreasing in importance. This indicates that any collinearity involved with these does not have a time component. The motivation, novelty, and valence terms had negative correlation coefficients in both conditions, especially in the Hamariyah condition which was far less noisy. This indicates that for some reason other factors may to increasingly overshadow these coefficients over time (i.e. other, collinear inputs to attention get higher β weights while theirs decrease). The factors that increase appear vary between the simulations, indicating that negative feedback may occur for certain factors. Negative feedback could occur due to the dynamics between factors. It could also occur due to feedback between a factor and attention. The correlation analysis in the next section will attempt to examine what relationships might exist that cause these trends in the β values.

7.3.4 Summary of Correlation Analysis of Simulation

Correlation matrices were generated to help examine which parameters covary. The full analysis is contained in Appendix J.1, since the full details are tangential to the core purpose of the internal validation. From examining these correlation matrices, it is clear that multi-collinearity is the reason that this approach inferring the salience weights ineffective. In both the Stanford Prison simulation and the Hamariyah simulation, a significant number of factors covary. As such, in a complex environment where such influences do not tend to be linearly independent, these weights cannot be easily inferred from the observable data.

From an empirical standpoint, this has interesting implications for attempting to determine the relative importance of these factors in affecting learning and attention. This analysis indicates that attempting to infer the interaction between such cognitive components using empirically collected data would require careful experiment design and should probably involve collecting data from multiple contexts. This is important because currently no experimental data gives a good idea on how these cognitive components truly interact. This analysis shows that even for a simple model, getting a good picture of such interactions would be difficult.

The correlation analysis also produced some interesting results with respect to which factors covary together in each simulation. From examining the correlations in Appendix J.1, two interesting correlations were found in both the Stanford Prison simulation data and the Hamariyah data. The first relationship was between novelty and attention, while the second was between valence and ingroup membership.

The first interesting pattern was found with how Novelty interacted with attention and the other components. In short, novelty correlates negatively with almost everything else. This relationship is extremely counter-intuitive at first glance. From the model design and initial analysis performed in Section 7.3.1, we know that increased novelty of an event increases its salience for attention. However, over the course of many events more novel events are less likely to be paid attention to.

Looking deeper into this pattern, it appears to be caused by a negative feedback between novelty and learning. Novel events are ones that are unfamiliar. Events can be unfamiliar (novel) because one has never been exposed to them, or because when exposed to them, one did not pay attention to them. This second reason for novelty is at the heart of the matter. A significant portion of novel events may remain unfamiliar because they were otherwise uninteresting. The negative feedback between attention and novelty occurs due to their shared relationship with learning. Attention is required in order to learn. Learning about an event decreases its novelty. So then, events salient mainly due to their novelty will lose novelty and be less salient. This means that even though novelty increases attentional salience, an event will only stay novel if the agent doesn't learn about it enough to become familiar with it. This means that persistent novelty of events indicates that those events are otherwise uninteresting. This is an interesting pattern, since it seems likely to occur in most adaptive systems-humans included.

The second interesting pattern was that membership in the same ingroup was correlated with higher valence. This makes intuitive sense, in that it would be expected that people sharing an ingroup would like each other more than people in outgroups. This dynamic matches with a significant body of research on social identity theories, which posit a preference toward people who appear to be in the same ingroup (Tajfel, 1982).

These two regularities may give some insight into empirical analysis of these cognitive phenomena. These are interesting and unexpected findings that show that the complete cognitive model provides some additional insight beyond the individual parts, an important part of any systems model. In particular, the relationship between novelty and attention appears to be a robust one. Finally, this analysis shows that it to be difficult to determine the relative impact of different factors that affect learning and attention. This analysis indicates that situational collinearity may cause certain factors to mask others, even to the point of some factors appearing to have a negative contribution even when it is known that they increase the likelihood of attention or learning.

7.3.5 Internal Validity Summary

The first analysis of the internal validity of the model showed that the computational model of agent cognition was implemented as intended. This shows that the model used by the simulations works as intended, from Chapter 5. This means that the computational model captures the empirically derived relationships used to design this model. It also showed that such relationships can be derived from the operation of agent cognitive model, by correctly estimating the relative attentional salience weights used by the model. This analysis confirms that the computational model works as expected.

The second analysis showed that a regression analysis could not derive these salience weights from either simulation data, due to situational collinearity between the factors. This was shown to not be an isolated effect, but one that should be expected in analyzing many contexts. In particular, it was notable that novelty will typically correlate negatively with attention if given enough time. This is due to the negative feedback between novelty and learning emerges in a complex environment (i.e. less attention → less learning → higher novelty). Secondly, membership in a common ingroup was observed correlate with higher valence. This may also be a typical relationship that would make it difficult to distinguish between the contributions from ingroup membership and valence (like/dislike).

Finally, many correlations between factors were different depending on the scenario. This result highlights the difficulty of inferring the relationships between different factors simultaneously. Since relationships between factors may be environmental rather than representative of cognitive factors, empirical studies on how these factors interact would need to be wary of how the experimental context can introduce collinearities. This is an important finding, since cognitive modeling can greatly benefit from broad scale studies that examine how large sets of factors interact. However, such experiments would only produce generalizable data for cognitive modeling if they avoid having high levels of situational connections between the factors being studied. The combination of the first and second internal validity analyses indicates that designing an experiment to improve upon this cognitive model design is theoretically possible, but great care would be required to develop appropriate experimental design.

7.4 Stanford Prison Experiment Simulation Analysis

The Stanford prison experiment produced a variety of results. The first results presented are the external validity measures, comparing the simulation results against the prison dynamics noted in the published papers and archival data. Next, a brief discussion of exploratory results is presented to examine how memes are transmitted in the social network.

7.4.1 External Validity Measures

Each external validity measure here compares data collected from the model against trends or data from the actual Stanford Prison Experiment. Table 7.17 lists the external validity measures that were applied to the Stanford Prison Experiment simulation and the ground truth empirical relationship. These measures were selected prior to observing the recorded data, and so can be used to determine how the recorded data corresponds with the ground truth. Each of these tests will be described briefly in the following section, and the results will be summarized at the end of the section in an analogous table.

Relative Action Proportions

The relative action frequencies are how often each action occurs, in comparison to other recorded actions. While the simulation was not intended to directly match the relative action frequencies, these frequencies were used to calibrate the first day of the experiment so some correspondence should be observed.

This analysis looked at the relative action frequencies in the simulation, among those that were coded from the Stanford Prison Experiment videos. As noted in Section 6.1.1, the following actions were implemented in the simulation and recorded: commands, information, insults, questions, resistance, physical aggression, helping, threats, use of instruments (threatening with a baton). The Stanford Prison experiment empirical data had counts of commands, helping, information, insults, resistance, and use of instruments. Due to reasons mentioned earlier (time discretization, background actions), it is impossible to directly compare raw counts. Instead, the raw count of each action is normalized by the total count of all these actions- generating the fraction of record actions that fall into each category. For the simulations, which had many runs, these

Table 7.17: External Validity Tests

Measure	**Empirical Ground Truth**
Relative Action Frequency	Examine how common actions are, compared to each other.
Command	Command was the most frequent action observed (≈ 38 % of actions)
Help	Help was the least common action observed (≈ 1%).
Information	Giving information was commonly observed (≈ 18% of actions)
Insult	Insults were commonly observed (≈ 18% of actions)
Resist	Prisoner resistance was somewhat common (≈ 9% of actions)
Threaten	Threats were somewhat common (≈ 9%)
Use of Instruments	Use of instruments was moderately common (≈ 8%)
Group Differences: Action Frequency	Examine which subsets of agents perform certain actions more than other agents.
Insult (S_13 vs Other Guards)	S_13 used more insults than other guards
Insult (Night Shift vs Other Shifts)	The night shift guards used more insults than guards in other shifts.
Command (Night Shift vs Other Shifts)	The night shift used more commands on prisoners than guards in other shifts.
Emotion Tests	Examine how emotions vary over time
Average Guard Emotions	Average guard emotions were slightly negative.
Average Prisoner Emotions	Average prisoner emotions were negative (about 3 times more negative than guards)
Guard Emotions Over Time	Guard emotions had a slight negative trend.
Prisoner Emotions Over Time	Prisoner emotions had a negative trend.
Meme Expression Ordering	Test which simulation condition best captures the order of first expression (if any).
ThrowInHole First Expression	Test which simulation condition(s) best predicts the order that guards throw their first prisoner in the hole.
Resistance First Expression	Test which simulation condition(s) best predicts the order that prisoners actively resist guards.
FeelImprisoned First Expression	Test which simulation condition(s) best predicts the order that prisoners feel that they are imprisoned and cannot escape.

fractions were averaged across runs. Table 7.18 shows the relative proportions of each type of action.

Table 7.18: Stanford Action Frequency Proportions (Mean)

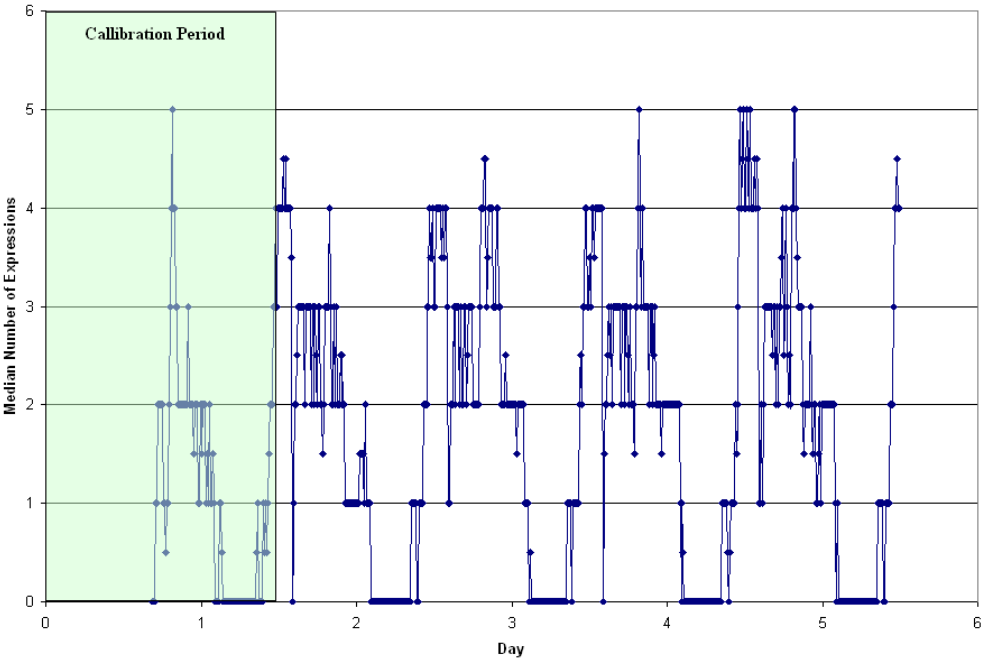
Action	**Ground Truth**	**Fully Known Condition**	**Hypothesis Condition**	**Authority Condition**
Command	0.38	0.27	0.29	0.29
Help	0.002	0.05	0.05	0.04
Information	0.18	0.16	0.15	0.16
Insult	0.18	0.05	0.03	0.03
Resist	0.09	0.29	0.30	0.30
Threat	0.09	0.02	0.01	0.01
Use Of Instruments	0.08	0.16	0.17	0.18

Looking at the data, it appears that despite the calibration over the initial portion of the experiment, the simulation runs showed significant deviations from the expected action distributions. Commands, Help, and Information each fall into their expected ranks- with Commands being a very common action, Information being somewhat common, and Help being uncommon. Commands, while still the most common guard action, were less common than in the actual experiment. Helping was slightly more common, but still very uncommon. Information showed almost an exact match. Insults were significantly less common in the simulation, as were threats. Instead, the use of instruments became a more popular action- picking up the slack for these. Since Use of Instruments, Threats, and Insults were functionally similar within the simulation, this is notable but not particularly interesting.

Resistance was significantly more common than in the actual experiment. To an extent, this was expected. During the actual experiment, a number of prisoners were dismissed early, while the simulation tended to retain the prisoners for the duration. The reasons for this difference are explained in Appendix H. The lower release rate resulted in more total prisoners, increasing opportunities for resistance. It is also possible that tuning based upon the first day made resistance more attractive than intended, since the original experiment showed little resistance on the first day. This may have caused resistance to be more common in the later portions of the experiment. Chart 7.3 supports this interpretation. Over the first 20 hours of the experiment that were used for calibration, the median resistance occurred at 1.31 resistance actions per time step. Over the full experiment, this value averaged 1.73 actions per time step- a 33% increase.

It appears that this is due to resistance occurring in a cyclic fashion. In particular, resistance is least prevalent during sleep periods and most common

Figure 7.3: Median Prisoner Resistance Over Time (Full Knowledge Condition)



during the count-off periods that follow meal times. It should be noted that different simulation runs showed slightly different trends in behaviors. In some simulation runs, resistance was minimal for the first three days of the simulation, while in others resistance was endemic across the experiment. If resistance in the actual experiment occurred in cycles, this would make exact frequency matching unlikely since only a small amount of the experiment was taped. Overall though, resistance is overexpressed within the model. This means that the later diffusion analysis for this meme will probably be significantly faster than might have occurred in the real-life situation.

Between-Group Action Proportions

The Stanford Prison Experiment states a number of empirical relationships found in action frequency, in terms of certain subgroups taking certain actions more than other subgroups. The simulation model should capture some of these dynamics, if it is appropriately modeling the situation. Three trends were noted: Guard S_13 used more Insults than other guards, the night shift (to which S_13 belonged) used more Insults than other shifts, and the night shift used more Command actions than other shifts.

Looking the Insult action, a secondary factor of interest was that S_13 tended to use insults more than other guards. Table 7.19 shows the mean number of expressions of this action in the simulation, for S_13 and the remaining agents. It also shows the p-value for a t-test that tested if the number of insults from S_13 was significantly higher than the average of other guards. An independent t-test was performed for each condition to test the probability that S_13 tended to use insults more. In the t-test, the first set of values was the number of insults performed by S_13 in each run and the second set of values was the average of other guards. The ground truth values are provided for reference, since they capture the relative proportions. However, these are not directly comparable to the simulation values since the ground truth are frequency rates while the simulation values are displayed as raw counts.

The simulation data, as shown in the table, indicate that S_13 used insults significantly more than other guards. In each of the experimental conditions, the harshest and most innovative of the guards, S_13, lived up to his John Wayne nickname in the simulation by showing a greater incidence of insults.

Table 7.19: Insult Frequency of S_13 vs Other Guards

Insult Metric	**Ground Truth (Rate)**	**Fully Known Condition**	**Hypothesis Condition**	**Authority Condition**
S_13 (Mean)	11.29	10.2	10.9	9.67
Other Guards (Mean)	∼0.99	6.1	2.84	2.15
T-Test Result	0.001	0.06	$6{\times}10^{-25}$	$9{\times}10^{-19}$

Table 7.20: Insult Frequency of Night Shift vs Other Guards

Insult Metric	**Ground Truth (Rate)**	**Fully Known Condition**	**Hypothesis Condition**	**Authority Condition**
Night Shift (Mean)	5.17	3.56	5.56	3.63
Other Guards (Mean)	2.29	7.72	2.83	2.59
T-Test Result	-	0.95	$7{\times}10^{-10}$	$5{\times}10^{-8}$

The Stanford materials also note that the night shift in general performed more insults. Table 7.20 shows the same analysis, comparing the night shift guards with the other guards. This shows a similar correlation, except for the Full Knowledge scenario. The Full Knowledge scenario had a modest increase in insults by guards outside the night shift and a decrease in insults by S_20 in the night shift. S_15, the final member of the night shift, was a “good guard” and did not tend to use insults much, as displayed by the large shift. This indicates that the night shift does not consistently evidence more insults, but instead that this dynamic hinges on the amount of insults used by S_20 who was the swing member on the shift. Digging deeper into the data, S_20 had much more variance in the Full Knowledge case than the other cases. Looking at the median level of insults by S_20 indicates that this low mean value is dragged down by outlier data, rather than being consistently different from the other cases. This indicates that the night shift did tend to use more insults than members of other shifts.

Table 7.21: Command Frequency of Night Shift vs Other Guards

Command Metric	**Ground Truth (Rate)**	**Fully Known Condition**	**Hypothesis Condition**	**Authority Condition**
Night Shift (Mean)	9.3	77.0	84.5	87.2
Other Guards (Mean)	4.04	73.0	82.08	82.0
T-Test Result	-	0.28	0.09	$1{\times}10^{-18}$

Commands tended to be higher for the night shift than other shifts, across all simulation conditions. However, this effect was not consistent in all conditions as shown by the poor results on the t-test for the Fully Known and Hypothesis conditions. While the night shift typically issues more commands in the simulation, it is not as pronounced as was seen in the empirical data.

The body of action data from the simulation correlates moderately well with that of the empirical data. While it is not fitting all values exactly, it captures most of the trends shown in the empirical data. It should be noted that unlike a statistical model or neural net, this cognitive modeling simulation is not prone to overfitting a scenario because a majority of the structural assumptions are based upon cognitive literature rather than the training data. From this perspective, the level of correspondence is reasonable.

Emotional Trends

The Stanford Prison experiment journal papers present some general emotional changes in prisoners and guards (Haney et al., 1973a, 1973b). In particular, prisoners had more negative affect over time while guards had only a minor decrease in affect. Since PMFServ uses emotions as part of its core framework, these are being used to test for this sort of trend. This analysis assumes that the Stanford Prison Experiment's empirical data about emotions should correspond to the aggregated emotions for agents in the simulated experiment, as defined in Section 7.1.

Using the Aggregate simulation emotions of each agent has some limitations compared to looking at the emotions individually. Firstly, it is possible that the trends of the aggregate are not representative of certain trends of the individual emotions. Secondly, it is possible that the Mood Adjective Checklist (MAC) trends are better represented by certain individual emotions rather than an aggregated emotional term. Unfortunately, the raw data from the MAC was not available so the only emotional trend data was derived from Haney et al. (1973a) and Haney et al. (1973b). These papers do not interpret the mood in detail and appear to be an aggregate emotional state of the subjects, hence the Aggregate emotional state of the simulated agents was used for comparison. The Aggregate emotional state is defined earlier in Equation 7.1 found in Section 7.1. Had more detailed ground truth information been available, each simulation emotion would have been analyzed and compared individually. This sort of analysis was not completed, since it would be difficult to interpret the results.

The first analysis was intended to test if the average values of agent emotions matched those from the empirical experiment. For each run, the average was calculated for agents in the Prisoner group and for agents in the Guard group. This was calculated to compare against the emotional trend data from the original Stanford Prison Experiment.

$$\overrightarrow{Emotions(Group)_r[t]} = \frac{1}{N}\sum_{x \in Group}^{N} Emotion(Agent_x)_r[t]) \tag{7.11}$$

To determine the emotional trends of each group, the emotions of the members had to be combined into a representative set of time series for the group. This was done by calculating the mean value of emotions for the group at each time point, for each run. This generated a vector of average emotions for a group for each run. Each element of the vector for any run r at a given time step t follows Equation 7.11. Table 7.22 shows the mean, median, and standard deviation for these values for each of the experimental conditions. It is evident in looking at the table that both guards and prisoners were somewhat unhappy in the experiment, on average. This matches the ground truth findings.

Table 7.22: Group Average Emotion Values

Group	**Mean**	**Median**	**Std Dev**
Guards (Full Knowledge)	-0.03	-0.05	0.05
Guards (Hypothesis)	-0.05	-0.05	0.03
Guards (Authority)	-0.05	-0.05	0.01
Prisoners (Full Knowledge)	-0.11	-0.13	0.05
Prisoners (Hypothesis)	-0.12	-0.12	0.03
Prisoners (Authority)	-0.13	-0.13	0.02

T-Tests were run on emotion data used to generate Table 7.22 to test if the guard emotions were higher than prisoner emotions, for each of the simulation conditions. In all conditions, the probability of the null hypothesis was p $<$ $1{\times}10^{-6}$. This strongly indicates that guards were happier than prisoners in the simulation. Comparing the means, prisoners were between 2.3 and 3.4 times less happy than the guards in the simulation. This corresponds well with the Stanford findings, which estimated prisoners as being about 3 times less happy than the guards. The average values of guards and prisoner emotions match with the ground truth values.

Table 7.23: Stanford Prison Simulation Aggregate Emotion Trends

Group	**Negative Trend %**
Guards (Fully Known)	7%
Guards (Hypothesis)	0%
Guards (Authority)	0%
Prisoners (Fully Known)	13%
Prisoners (Hypothesis)	10%
Prisoners (Authority)	3%

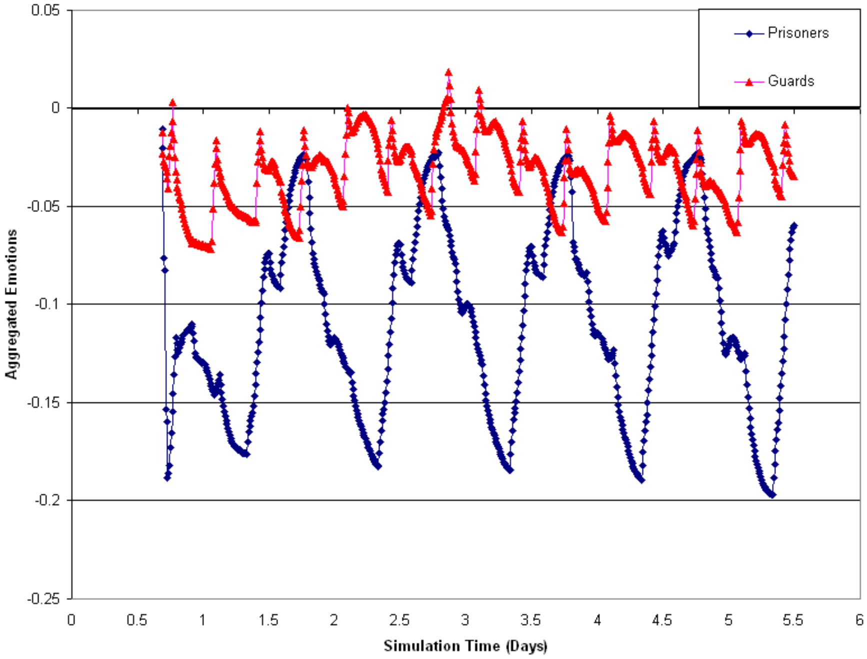
Emotions were also expected to get worse over time, particularly for the prisoners. For each run, the group of agents of interest (guards or prisoners currently in the experiment) had their emotions averaged and entered as a time series. A Mann-Kendall test was used on each group's time series, for each run. Table 7.23 lists the number of runs in each experimental condition where the Mann-Kendall test indicated that emotions decreased with $p < 0.05$.

This Mann-Kendall trend analysis strongly contradicts the expected result from the empirical results. Emotions did not steadily worsen in the simulation but were volatile and non-monotonic. For example, Figure 7.4 plots the mean value of aggregate group emotions for prisoners and guards. Emotions for individual subjects had similar patterns, where certain periods made them less happy than others. The trends were not entirely time dependent, but in most cases some level of emotional cycling occurred.

This means that the emotions in the PMFServ simulation cannot be adequately compared with the Stanford Data. Since emotions in the Stanford Prison experiment were only calculated in three (incomplete) point samples, they provide limited correspondence information for comparing against a cyclic time series. Even if the real life situation fit these curves precisely, any three equidistant samples could produce a positive or negative trend. For this reason, the trend analysis is inconclusive.

The validation based upon emotions showed mixed results. While guard and prisoner emotions had appropriate average values, they appeared to work cyclically rather than in monotonic trends. However, the ground truth stated in Haney et al. (1973a) do not state that the trends are monotonic but explicitly states that prisoners had significant emotional volatility. As such, it appears there is insufficient ground truth data to draw any strong conclusions about validity

Figure 7.4: Aggregate Group Emotions (Full Knowledge Condition)



on this aspect.

Meme Expression Ordering

The meme expression ordering is the most important external validity test. As noted in Section 6.1.2, the order of first expression for three different memes was inferred from the original data sources. The first analysis performed was to calculate the median first expression orderings for agents expressing each meme. Table 7.24 shows the ground truth orderings next to the median orderings from the simulation runs under each condition. Tables 7.25 and 7.26 show these same results for the Resist meme and FeelImprisoned meme respectively. As noted previously, agents shown in parenthesis typically did not express the meme within the experiment.

The orderings for each condition have significant similarities, both to the ground truth condition and each other. S_11, S_12, S_13, and S_20 each tended to express earlier than other agents. S_16 and S_19 tended to express later than other agents. To an extent, this is influenced by the shift ordering. Following a brief day shift, the night shift (S_13, S_15, S_20) takes over. Even when all guards are aware of their ability to throw a prisoner in the hole, this does not tend to happen until the evening shift takes over. However, the hypothesis condition

Table 7.24: Stanford First Meme Expression Ordering Results (Throw In Hole)

Ground Truth	Full Knowledge	Hypothesis	Authority
S_13	S_13, S_20	S_13	S_13, S_15
S_20	S_15	S_20	S_12
S_11	S_12	S_11	S_11
S_12, S_18	S_11	S_12	S_18
S_16, S_17, S_21	S_18	S_17	S_17
(S_15, S_19)	S_17, S_21	S_15	S_21
	(S_16, S_19)	S_21	S_20
		S_18	S_16
		(S_16, S_19)	(S_19,)

matches the ground truth slightly better- accurately reflecting the first 4 elements of the sequence, and showing S_15 as later in the sequence rather than earlier. S_15 was the "good guard" on the evening shift, who did not tend to imitate S_13.

For this ordering, it is evident that the Full Knowledge condition performs fairly well, but that the hypothesis condition may be capturing a key interaction that allows S_15 to express the meme later. The Authority condition has non-intuitive outcomes for the first shift, including the non-intuitive issues of S_15 sometimes originating the meme and S_20 being one of the last adopters. Otherwise, the Authority condition is very similar to the Full Knowledge condition. Looking deeper at the data, this similarity is caused by a majority of guards learning about the meme if it is presented to them by a figure of high authority. Thus, the Authority condition typically starts with a majority of agents aware of their respective memes. The median-value analysis seems to slightly favor the Hypothesis condition, but Full Knowledge also gives a reasonable median ordering.

Table 7.25: Stanford First Meme Expression Ordering Results (Resist)

Ground Truth	Full Knowledge	Hypothesis	Authority
S_05	S_01	S_05	S_01
S_09	S_06, S_09	S_01	S_06
S_01, S_04	S_04	S_03, S_04	S_09
S_06	S_03	S_06, S_08	S_04
S_08	S_08	S_09	S_03
S_03	S_05	S_00	S_08
S_00	S_02	(S_02,)	S_05
S_02	S_00		S_00
			S_02

As shown in Table 7.26, the Resist orderings show some similar trends. The Full Knowledge and Authority conditions are more similar to each other than to the Hypothesis condition. A notable difference is that the Hypothesis condition is the only condition where S_05 is the first to express the Resist action. Full Knowledge and Authority conditions place S_05 as expressing much later. Secondly, the Hypothesis condition shows far more variance in its ordering toward the middle (positions 3-7). S_03, S_04, S_06, and S_08 each have similar median orderings that vary from run to run. Finally, the Hypothesis condition predicts S_09 as being far later than either simulation condition or the ground truth. No condition clearly outperforms another in the median value analysis of resistance. The Hypothesis condition performs well, but it misses completely on S_09 and shows uncertainty about the middle. The Full Knowledge condition and Authority conditions are more certain in their orderings and place each agent close to its appropriate ordering, but miss on S_05 who was one of the notable resisters. Moreover while most orderings are close, the median condition for these does not resemble the exact ordering of the ground truth.

Table 7.26: Stanford First Meme Expression Ordering Results (Feel Imprisoned)

Ground Truth	**Full Knowledge**	**Hypothesis**	**Authority**
S_05	S_01	S_01	S_01
S_02, S_03	S_05, S_06, S_09	S_05, S_06	S_03, S_05, S_06, S_09
S_06	S_02, S_03, S_04	S_02, S_09	S_02, S_04
S_01	S_00	S_03	S_00
S_09	(S_08)	S_04	(S_08)
S_00		S_00	
(S_04, S_08)		(S_08)	

For the FeelImprisoned meme, all experimental conditions show a similar ordering for the median condition. This is expected, since the FeelImprisoned meme had the same starting set of agents aware of the meme across all conditions. As such, this consistency between conditions is expected. All conditions show high variability toward the middle and indicates that S_01 expresses feeling imprisoned much earlier than the ground truth condition. In general, the FeelImprisoned meme does not appear to be captured as accurately as the other memes.

$$f_{Inv} = 1 - \frac{Inversions}{MaxInversions} \tag{7.12}$$

To look at this from a different perspective, the inversion distance was calculated between simulation run orderings and the ground truth orderings. For each run in each condition, the inversion count algorithm was run- calculating the number of inversions and the maximum number of inversions, given the observed sequence. This nearness is defined in Equation 7.12, where Inversions is the distance metric defined in Section 7.2.4 and MaxInversions is the calculated value of the farthest the sequences could be apart. For each action under each experimental condition, the average nearness was calculated across the set of runs in that condition. This gives a metric for how close the ordering was the actual order, with a value of 1 being a perfect match and a value of 0 being the worst possible match. As a result, a higher nearness value is better.

Table 7.27: Stanford First Meme Expression Order Nearness

	Full Knowledge		**Hypothesis**		**Authority**	
Meme	Avg(f_{Inv})	StdErr(f_{Inv})	Avg(f_{Inv})	StdErr(f_{Inv})	Avg(f_{Inv})	StdErr(f_{Inv})
ThrowInHole	0.79	0.013	0.66	0.017	0.78	0.017
Resist	0.71	0.009	0.75	0.015	0.70	0.008
FeelImprisoned	0.64	0.018	0.65	0.017	0.68	0.014

Table 7.28: Stanford First Meme Expression Nearness: % f_{Inv} >0.5

Meme	**Full Knowledge**	**Hypothesis**	**Authority**
ThrowInHole	100%	100%	100%
Resist	100%	100%	100%
FeelImprisoned	87%	97%	100%

Table 7.27 lists the average nearness of the 30 simulated orderings to the ground truth ordering, along with the standard error term. Table 7.28 displays the percent of runs which do better than chance (>0.5) for each meme, to display the consistency that runs performed better than chance. Tables 7.29 and 7.30 show these same statistics, but ignores the first element from the ground truth sequences. This is because the first element gives the hypothesis condition an advantage, since that agent must be the first one to express the meme (since they are the only one to start with it). This means that the innovators noted in the prior sections are not considered in this analysis. For ThrowInHole, this involved removing S_13 from analysis, for Resist S_00 and S_05 were removed, and for FeelImprisoned S_08 was removed.

Looking at the ordering analysis, the first conclusion that can be drawn is that all conditions perform significantly better than chance. These results are consistent, with the ordering performing better than a random sequence at rates ranging between 57% and 100%. In particular when using the full orderings, Resist and ThrowInHole perform better than chance in 100% of the runs. As was evident in looking at the median orderings, the expression orderings for

Table 7.29: Stanford First Meme Expression Order Nearness (No Innovators)

	Full Knowledge		**Hypothesis**		**Authority**	
Meme	Avg(f_{Inv})	StdErr(f_{Inv})	Avg(f_{Inv})	StdErr(f_{Inv})	Avg(f_{Inv})	StdErr(f_{Inv})
ThrowInHole	0.74	0.015	0.54	0.023	0.74	0.020
Resist	0.77	0.016	0.55	0.026	0.77	0.012
FeelImprisoned	0.54	0.022	0.55	0.021	0.58	0.017

Table 7.30: Stanford First Meme Expression Nearness: % f_{Inv} >0.5 (No Innovators)

Meme	**Full Knowledge**	**Hypothesis**	**Authority**
ThrowInHole	100%	57%	100%
Resist	100%	57%	100%
FeelImprisoned	67%	60%	77%

FeelImprisoned were less accurate while ThrowInHole and Resist were fairly accurate.

Running the inversion analysis on the median orderings shows even better correspondence, with the Hypothesis condition and the Full Knowledge condition there are marked improvements in the correspondence. By using the median ordering to obtain a typical run, correspondences are higher than the average across individual runs. Table 7.31 shows the correspondence of the median sequences to the ground truth sequences, with and without innovators. This analysis also indicates that the Full Knowledge and Hypothesis conditions appear to perform better than the Authority condition.

Table 7.31: Stanford First Meme Expression Inversion Results (Median Sequences)

	Full Knowledge		**Hypothesis**		**Authority**	
Meme	All	No S_13	All	No S_05 or No S_00	All	No S_08
ThrowInHole	0.82	0.77	0.85	0.81	0.67	0.58
Resist	0.71	0.84	0.79	0.61	0.69	0.80
FeelImprisoned	0.69	0.59	0.75	0.68	0.73	0.63

Looking at the totality of the ordering analyses, the Full Knowledge condition appears to have the best overall performance. For ThrowInHole, the Full Knowledge condition performs well on the individual sequences and on the median-ordered sequence. The Hypothesis condition works very well on the median-ordered sequence, but it performs worse on the individual sequences than Full Knowledge. The Authority condition performs as well as the Full Knowledge condition, at best.

The Resist orderings are more complicated. The Hypothesis condition performs the best when looking at the full orderings, with modestly better correspondence than Full Knowledge. However, removing S_00 and S_05 pushes the Hypothesis condition down to barely better than chance. Again, this reinforces the median-ordering analysis which showed the Hypothesis condition to perform a bit worse toward the middle. Correspondingly, the Full Knowledge condition performed worse on S_05 and improves when that subject removed from analysis. Each condition performs similarly on the FeelImprisoned ordering and shows a modest correspondence that is better than chance.

The Full Knowledge worked reliably on both Resist and ThrowInHole potential memes. For the ThrowInHole meme, it appears possible that all agents knew about their ability to throw a prisoner in the hole at the start. This means that their orderings were probably determined by the nature of how they interacted with the prisoners, their personality differences, and the ordering of guard shifts. Additionally, the Full Knowledge condition is more reliable for representing the ordering of resistance among prisoners. This makes Full Knowledge a plausible condition.

The Hypothesis condition's strong performance on the median-orderings supports it as a plausible mechanism in the Stanford Prison. The Hypothesis condition performed worse than the Full Knowledge condition on average, but was very effective on the median-ordering analysis. Since the median ordering provides a "typical run" and on the typical runs, the Hypothesis condition has some advantages over Full Knowledge. However, on any particular run the Full Knowledge condition appears to capture the orderings better.

The Authority condition was not as plausible as the other conditions for simulating the order of first expression. Authority tends to perform similarly to Full Knowledge, but slightly worse. Looking at the underlying data, the Authority condition typically starts with a majority of the agents aware of each meme. This means that limited learning could take place, and the transmissions that did take place did not seem to improve the correspondence. As such, the Authority condition is nominally plausible but seems less plausible than Full Knowledge.

With that said, the Authority condition only tested the impact of a generic authority figure as opposed to any specific individual. While this Authority condition is less plausible, it is still possible that social learning of ThrowInHole was the result of guards interacting with the prison warden or due to social factors other than Authority alone. Unfortunately, these conditions are not testable since it is both infeasible and inappropriate to model the experimenters for involved in the situation. This means that the Authority condition cannot be tested further using this model and should be considered less plausible than the other conditions.

In conclusion, both the Hypothesis and Full Knowledge conditions are both plausible and this simulation does not definitively show that one condition was more probable than the other. Both perform well on different metrics, but it is unclear which statistic has a better predictive power because this is a newly implemented model. Finally, it is also possible that a mixed condition existed, where social learning and imitation occurred differently than the explored conditions. For example, a subset of guards might have discussed the ThrowInHole action during orientation. This might explain why Hypothesis and Full Knowledge appear to work in complementary ways.

External Validity Metrics Summary

As has been shown, a number of external validity tests were applied to examine if the Stanford Prison simulation captures important dynamics from the empirical study. Table 7.32 summarizes these tests, broken down by the performance of the model on them under each condition. Each correspondence is rated as: Hit, Close, Poor, Miss, or Unclear. Hits are defined as validity results that appeared to capture the correct relationships and relative values. Close is used to rate correspondences that did not capture the exact value, but they still capture important relationships (such as ordinal comparisons). Poor correspondences are validity metrics performed better than chance but are not very close to the right values. Misses are tests that firmly assert a value that contradicts the ground truth. Unclear tests are ones that could not be effectively evaluated, due to ambiguous ground truth data.

Looking at the results, a few things are notable. Firstly, the simulation does fairly well on the correspondence tests. There are a large number of metrics that either provide near-exact matches or closely resemble the ground truth data. While there are some misses, these situations are explainable within the context of the model and could have been addressed if more training data was used to tune the simulation. Secondly, the different simulation conditions perform fairly similarly on most metrics. While there are some small changes, the only major differences appear in the meme transmission. This indicates that the simulation is fairly robust with respect to these different conditions.

The ultimate intention of this experiment was to examine the possibility that memes might have been a factor in how events unfolded in the Stanford Prison Experiment. Unfortunately, this analysis did not conclusively resolve this question. Qualitatively from the median-ordering analysis, memes did seem to provide a better match to the expected behavior. However, the individual runs in the Full Knowledge condition had a better match with the correct orderings. The important deciding factor between these measures

Table 7.32: External Validity Test Results

Measure	**Full Knowledge**	**Hypothesis**	**Authority**
Relative Action Frequency			
Command (0.39)	Close	Close	Close
Help (0.01)	Close	Close	Close
Information (0.18)	Hit	Hit	Hit
Insult (0.18)	Poor	Poor	Poor
Resist (0.09)	Miss	Miss	Miss
Threaten (0.09)	Miss	Miss	Miss
Use of Instruments (0.08)	Close	Close	Close
Group Differences: Action Frequency			
Insult (S_13 > Other Guards)	Close	Hit	Hit
Insult (Night Shift > Other Shifts)	Miss	Hit	Hit
Command (Night Shift > Other Shifts)	Poor	Close	Close
Emotion Tests			
Average Guard Emotions < 0	Hit	Hit	Hit
Average Prisoner Emotions < 0	Hit	Hit	Hit
Average Prisoner Emotions ≈ 3×Avg Guard Emotions =	Close	Close	Close
Guard Emotions Have Negative Slope	Unclear	Unclear	Unclear
Prisoner Emotions Over Time	Unclear	Unclear	Unclear
Meme Expression Ordering			
ThrowInHole First Expression	Close	Close	Close
Resistance First Expression	Close	Close	Close
FeelImprisoned First Expression	Poor	Poor	Poor
Verification Summary			
Hit or Close	10	12	12
Miss	3	2	2

would be their predictive value: Is the median-ordering of multiple agent-based experiments a better predictor of real-world events than the correspondence of individual run orderings? Unfortunately, since this is a new model that question remains unanswered. The result of this analysis indicates that memes are a plausible mechanism for the progression of these behaviors in the Stanford Prison Experiment, but that full knowledge of these actions was equally plausible.

Finally, while it is possible to test that meme transmission may have been a mechanism in the Stanford Prison Experiment, this analysis cannot validate the content of the memes. While there might have been a meme for knowing the action of how to throw a prisoner in the hole, there could instead have been a meme which made guards aware that no one would punish them if they threw a prisoner in the hole. As is typical with studying learning, the exact content of the learning can be hard to infer. As such, while memes have been shown to be potentially plausible- this external validity testing cannot speak to the content of memes. It has simply showed that memes related to these actions can be a plausible mechanism for looking at the Stanford Prison Experiment.

7.4.2 Exploratory Analysis

The exploratory analysis for the Stanford Prison Experiment was intended to examine two issues. First, an examination of the diffusion dynamics was conducted. As noted in Section 7.2.5, it is valuable to know the rate that different memes diffuse through the population. A second analysis was conducted to determine the transmission dynamics, in particular which agents tended to learn about a meme from which other agents. This section will focus on the Hypothesis condition, as it demonstrated most diffusion of memes.

Diffusion Dynamics

The diffusion dynamics memes can be looked at from two standpoints: learning and expression. Within the simulation, different memes have different diffusion rates and different amounts of delay before they are expressed by other agents. Figures 7.5 and 7.6 show the number of agents aware of the Resist and ThrowInHole memes over time respectively, for the Hypothesis condition. The thick central line of these figures indicates the mean value, while the dashed lines show the progression from individual runs. It should be noted that these figures show learning by all agents, not just the ones able to express the memes.

Figure 7.5: ThrowInHole Learning Diffusion

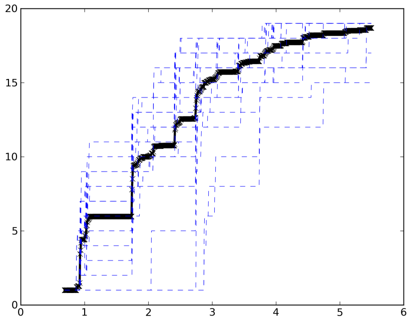
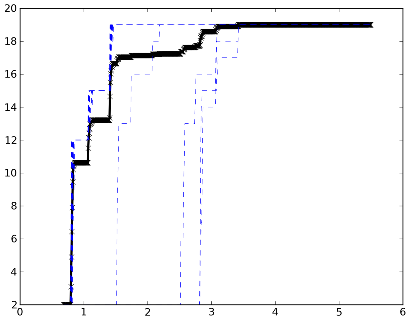


Figure 7.6: Resistance Learning Diffusion



Both curves can be approximated using s-curve form, as expected. Additionally, the individual runs tended to be similar to the average run with only a few deviant outliers. This was typical for diffusion results using this cognitive architecture in general. Looking at the learning-diffusion charts, it is evident that ThrowInHole had a slower diffusion rate. Logistic curves were fit to the mean-value curve in each of these graphs, which confirm this interpretation. SciPy, a scientific package for Python, was used to fit generalized logistic curves to each mean-value curve. The learning rate for ThrowInHole was 1.37, while the learning rate for Resist was 3.45. This is approximately double. A major factor in this is that Resist occurred more often than ThrowInHole. This happened for two reasons. Firstly, ThrowInHole was an action that placed a prisoner in the Hole until they were released. With the problem prisoner removed, there was less need to express the meme. Secondly, Resist occurred more often than expected by the ground-truth for reasons as previously explained in Section 7.4.1.

Figure 7.7: ThrowInHole Learning Diffusion (Guards)

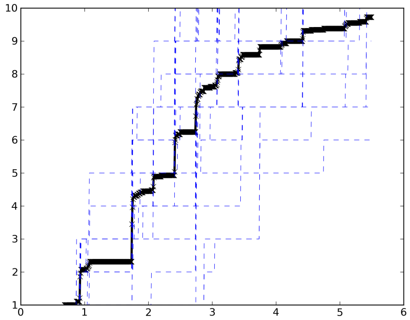
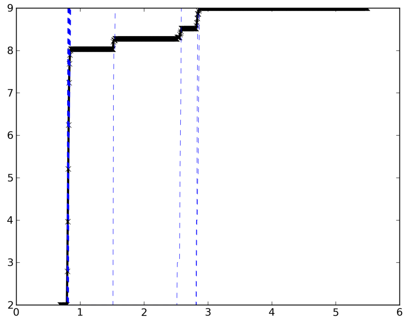


Figure 7.8: Resist Learning Diffusion (Prisoners)



This difference in learning rate is greatly increased if one looks only at the agents who can express each meme. Since guards worked in shifts, there are distinct periods where information is exchanged. By comparison, the prisoners were in a common environment for the entire simulation. Figure 7.7 shows the number of guards who are aware of ThrowInHole over time in the Hypothesis condition, while 7.8 shows the number of prisoners aware of the Resist action over time. The learning rates for guards in this case are 1.45, while the prisoner learning rate is 10.8. Moreover, the graphs show that the learning was qualitatively different. The individual runs for guard learning show that learning tended to happen mainly at shift transitions, but also at sporadic points during shifts. Conversely, prisoner learning tended to occur in sharp bursts. Once three or more prisoners became aware of the Resist meme, the rest of the prisoners became aware of it within hours during the simulation.

Having looked at the learning rates, it is logical to examine how these relate to the number of agents who have expressed the meme at least once. Figure 7.10 shows the counts of first expressions over time for prisoners taking the Resist action. Interestingly, while prisoner learning diffuses quickly, this does not necessarily correlate to all prisoners resisting. While at least five prisoners tend to

Figure 7.9: ThrowInHole First Expression Diffusion

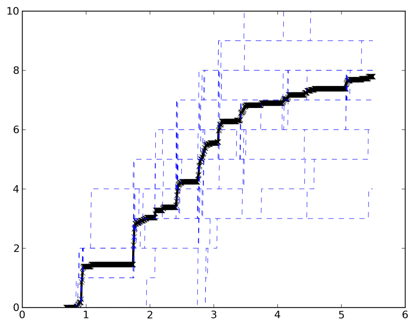
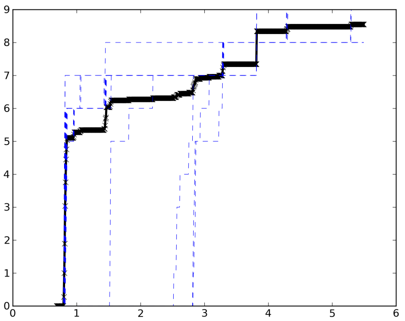


Figure 7.10: Resist First Expression Diffusion



express resistance fairly quickly after learning it, not all prisoners use resistance and the times at which they first decide to resist are variable. Figure 7.9 shows the counts of first expressions over time for guards taking the ThrowInHole action for the Hypothesis condition. Guard first expression times correlate strongly with shift changes, as was seen with learning. This indicates that learning related to the ThrowInHole action could have been an influence for when particular guards started throwing prisoners in the Hole.

Figure 7.11: ThrowInHole Learning vs First Expressions

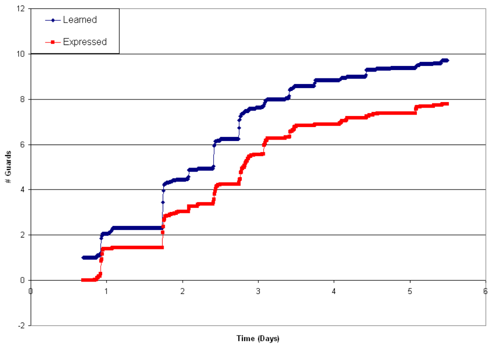
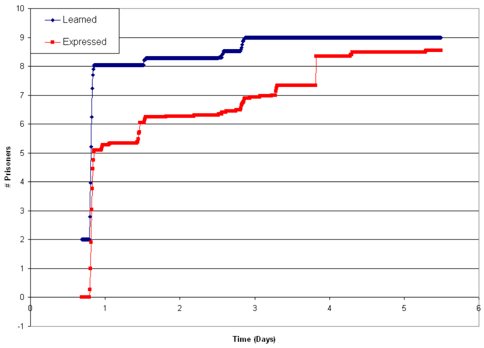


Figure 7.12: Resist Learning and First Expressions



Figures 7.11 and 7.12 compare the mean values for learning and first expression counts for ThrowInHole and Resist, respectively. From these, it is clear that the first expression times for agents correlate with the learning times. However, these graphs do not establish how much each one causes the other. Learning the meme is a causal factor for expressing it. However, an agent expressing a meme for a first time may help it reach agents who previously were not paying attention to it. In this way, an agent's first expression can be a causal

factor for learning.

To test for this, Granger causality tests were applied to the time-series of learning and first expression sequences for each run. The tests were run in both directions, to examine which sequence showed better causality for the other. Since Granger causality tests are sensitive to the lag parameter, each test was iterated over a sweep of lag values between 1 and 6 (one hour of simulation time) to find the optimal lag for each test. Table 7.33 shows the results of this analysis for the Hypothesis condition. The percent of runs where the causality test was significant (p < 0.001) is indicated for each meme, as well as the average lag time which was optimal in those significant cases.

Table 7.33: Causality: First Learning vs First Expression

	Learning		**First Expression**	
Meme	% p-value<0.001	Avg(Lag)	% p-value<0.001	Avg(Lag)
ThrowInHole	97%	3.96 (39.6 min)	21%	3.5 (35 min)
Resist	100%	3.66 (36.6 min)	86%	3.32 (33.2 min)

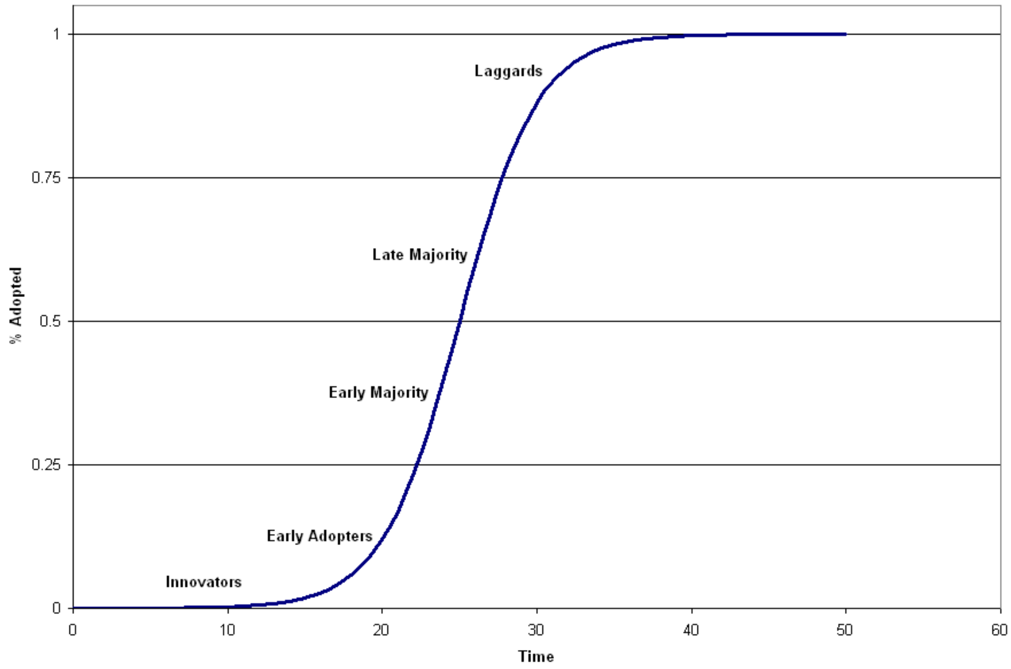
This analysis indicates that in both cases, first learning appears causal to first expression. This is as expected, since learning is necessary for first expression. The lag indicates that typically agents tend to use the action within 40 minutes after learning it. However, when using a more full sweep with lags up to 1/3 the length of the simulation, optimal lags as long as a full day were found. These might indicate cases where agents tended to learn the ThrowInHole action, but did not take it until their next guard shift.

First expressions are also causal for first learning for Resist, but seldom for ThrowInHole. This may indicate that ThrowInHole was transmitted mainly by a subset of agents, whose expression was a key causal factor for learning. Resistance did not show this trend. This indicates that the first expression of resistance for an agent was a more strongly causal factor for learning. This could mean that first expression is more likely to reach new agents for learning, but it may also result from first expression correlating with the number of expressions in general. To examine this, it is important to look at how agents learned memes, which is the focus of the next section.

Meme Transmission Dynamics

This section explores the who and why of meme transmission. This is an interesting feature of the model, since it allows detailed analysis of who was spreading memes and when they were most effective. The question of who was expressing memes will be examined by breaking down the agents into classes based on their tendencies to learn and express the meme. Using these classes, the question of why is examined by looking at the differences in personality factors between agents in different classes.

Figure 7.13: Typical Adoption Curve Positions



Rogers (1962) separates adopters into 6 categories: innovators, early adopters, early majority, late majority, and laggards. These indicate the different phases of adoption on the s-curve, as shown in Figure 7.13. Due to the low number of agents, early adopters and early majority will be lumped together. The first analysis performed was a quartile ranking of agents' relationship to the meme that they could express (Resist or ThrowInHole). Quartiles were calculated for the following metrics: the average time an agent first learned the meme, the average time they first expressed it, the average number of exposures they took to learn it, and the average fractions of their actions that expressed the meme once it was known. Note that for the number of exposures to learn a meme, multiple agents act simultaneously so multiple exposures can occur in the same step. As such, all exposures in the step where they learn are counted, including the ones they learn the meme from. The full tables of these quartiles are contained in Appendix J.3. These helped provide the insight for the following analysis.

The first learning and first expression times were used to examine which agents could be considered the early adopters versus the laggards. Table 7.34 shows this information for the Resist action under the Hypothesis condition. For

Resist, prisoners all tended to learn quite quickly- on average less than an hour apart. They were effectively all early learners. However, expression was phased out much differently. Certain agents such as S_01 and S_04 were much quicker to resist once it was demonstrated to them. Conversely, agent S_02 did not tend to resist until much later, if at all. This is notable, since S_02 was the agent whose strategy was to vigorously go along with the guards.

Table 7.34: Resist Adopter Categories

	Learning			
Expression	**Innovators**	**Early**	**Late**	**Laggard**
Innovators	S_05			
Early		S_01, S_04, S_06, S_08		
Late		S_03, S_09		
Laggard	S_00	S_02		

Table 7.35: ThrowInHole Adopter Categories

	Learning			
Expression	**Innovators**	**Early**	**Late**	**Laggard**
Innovators	S_13 (E)			
Early		S_15 (E), S_20 (E)	S_17 (D), S_21 (D)	
Late				S_11 (N), S_12 (N)
Laggard		S_16 (D), S_19 (D)		S_18 (N)

The ThrowInHole meme worked very differently, as seen in Table 7.35. Due to the shift boundaries and attention issues, the meme rolled out in a much more staged fashion. To show the effect of shifts, each guard is followed by their shift: Day (D), Evening (E), or Night (N). S_13, the innovator, was part of the evening shift. S_16, S_17, S_19, and S_20 appear to learn during their first cross-over period, where they are leaving their shift and S_13 is starting. One interesting aspect of this is that the more pacifist agents, S_16 and S_19, learned the meme before the ones who expressed the action earlier, S_17 and S_20. To a lesser effect, this was also seen with S_18 versus the other evening shift guards. Guards who were less likely to use the ThrowInHole action were quicker to attend to it. This is an interesting and counter-intuitive effect.

A second question of interest is the matter of who are resistant to the meme and who are more likely to express the meme once they know it. These two factors are interesting to look at together because they show who will tend to be the expressive early adopters, passive carriers, resistant but later expressive, or holdouts. By knowing these factors and the network topology, it would be possible to get a good estimate of the diffusion rate: resistance shows the number of exposures needed to acquire the meme, while the expression rate gives the output of exposures. The number of exposures required to learn the meme was considered to be a resistance factor- more exposures indicates the agent was less susceptible to the meme in this condition. The intermediate processing of this is contained in Appendix J.3 also.

Table 7.36: ThrowInHole Meme Resistance vs Expressiveness

	Expression			
Resistance	Highest	Higher	Lower	Lowest
Lowest	S_11		S_13	
Lower		S_12	S_15	S_19
Higher	S_20		S_18	
Highest	S_21	S_17		S_16

Table 7.36 shows each guard agent's resistance (# exposures to learn) and its expressiveness (fraction of actions producing the meme) for the ThrowInHole meme. This analysis reproduces much of what had been expected. S_15 and S_19 ("nice" guards) learn the meme readily, but don't express it very often. On the converse, guards such as S_21 take some additional exposures before learning the meme but regularly throw prisoners in the Hole once they learn it.

The same analysis was applied to the Resist action and the prisoners, with the results as shown in Table 7.37. Unfortunately, while the ThrowInHole results were quite reasonable- the Resist results show some of the issues that were present in the internal validity checks: the expression of the Resist action happens a bit too often, but does not occur quite often among the innovators. S_05 and S_00, while expressing the meme at a reasonable rate, do not express it quite as high as some other agents. This may indicate that the simulation did not capture a factor beyond their basic personality traits led to increased Resist actions from these agents. Both agents had an ideological background, with S_00 believing in meditation and S_05 supporting Marxist-type ideology. In other respects, the classifications seem reasonable. S_02 is shown to be the holdout, both for learning the meme and expressing it. Agents that were known to use Resist with some frequency, such S_04 show up as more expressive than other agents.

At first glance, it seems as if the early adopters take more exposures to learn the meme, on average. To look into this further, a metric was devised to look at the fraction of exposures that lead to learning as a function of the number of agents aware of the meme. This metric only counts new learning, and any exposures on agents with the meme are ignored. This gives an estimate of the exposure efficiency, its ability to cause new learning. Figure 7.14 plots the exposure efficiency of ThrowInHole on guards as a function of the number of

Table 7.37: Resist Meme Resistance vs Expressiveness

	Expression			
Resistance	Highest	Higher	Lower	Lowest
Lowest			S_00	S_05
Lower		S_03, S_06		
Higher	S_04, S_08	S_01		
Highest			S_09	S_02

agents knowing the meme. Figure 7.15 shows exposure efficiency for the Resist action on prisoners. In these charts the solid line represents the mean of the points, while the dashed line shows a fitted 2nd-order polynomial. This analysis indicates that the first learners are not disadvantaged with respect to salience. This is supported by looking at the median resistance levels, where the early adopters show lower resistance rather than higher resistance. The reason why they take more exposures to learn, on average, is due to having longer tail distributions. While later adopters may get multiple simultaneous exposures, helping to smooth the distribution, the early adopters generally only see one exposure at a time.

Figure 7.14: Exposure Efficiency for ThrowInHole

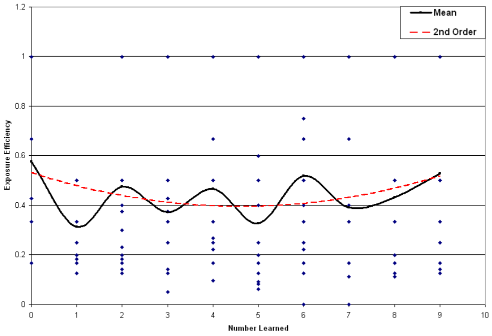
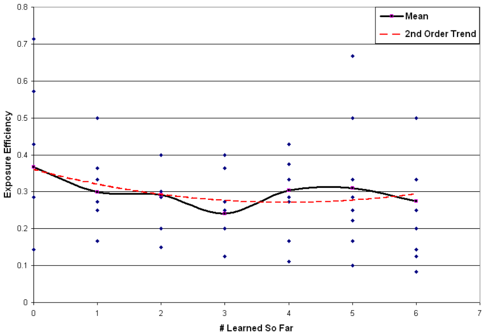


Figure 7.15: Exposure Efficiency for Resist



The efficiency curves are interesting in their own right. Both figures show a slight U-curve, where efficiency is lower between the early adopters and the early majority. This is an interesting effect, since it implies that diffusion occurs slightly differently than a traditional diffusion curve. This may indicate that in a social environment memes may tend to have an initial growth spurt, followed by a lull. This makes some intuitive sense- the most interested parties will pay attention early. Despite this, the overall diffusion rate is faster as more agents learn because there are typically more total expressions- even if they are less effective per expression. This is an interesting lead, which may be worth looking into in future experiments.

Stanford Prison Experiment Summary

Overall, the Stanford Prison Experiment simulation has provided a number of interesting avenues for future research. Firstly, it has demonstrated an effective simulation of a real-life scenario. Secondly, it has demonstrated the ability to extend classical diffusion of innovation simulation. This simulation takes into account both physical and social environments, combining social influence and physical barriers (such as shift change) into a common framework. Thirdly, this simulation has shown unique capabilities beyond typical diffusion-of-ideas research. This fine-grained analysis allowed identifying the different phases of adoption for individuals, as well as to determine their relative level of output value. This is very different from a classical diffusion model, which seldom models the background actions that agents can take instead of expressing a meme.

Finally, it has opened new avenues of simulation and empirical research. For example, the internal validity analysis showed that novelty correlates negatively with attention in a complex environment. While initially counter-intuitive, this is a useful finding that appears likely to be reproducible within an experimental setting. The internal validity analysis on attention also highlighted the ability of an agent-based model to help explore how situational factors could affect trends in data. This demonstrated the ability of the model to be a test-bed for mocking up an experimental condition.

7.5 Iraqi Village Simulation Analysis

The Hamariyah Iraqi Village simulation was intended to examine a much more focused issue that the Stanford scenario could not capture: meme competition. While the Stanford prison simulation had two defined groups who were only able to express one of the memes, the Hamariyah scenario allows all agents to perform both memes: Give Information and Plant IED. The first meme involves going to a US-owned building to provide a tip about insurgent activity. The second meme involves volunteering to plant an IED on a US-owned building. As noted in Section 6.2.2, the meme actions were calibrated to help distinguish between agents who preferred only one of the memes, liked both memes, or disliked both memes. This allows for cross-class comparison to determine the personality and situational characteristics common to these groups.

7.5.1 Diffusion Dynamics

The simulation dynamics give an overview of how the memes progressed. Both memes progressed faster than would be expected in a real-life scenario, for learning and expression. This was due to modeling choices noted in Section 6.2.2, where the model was designed to discriminate which memes agents were likely to learn and express, rather than intending to model the base expression rate accurately. As such, this section differs slightly from the previous examination of diffusion. While the prior section showed raw diffusion rates, this one will attempt to explore the differences in relative rates of learning and expression.

As noted in Section 6.2.4, the Hamariyah village was run under two different starting conditions: random agents aware of the memes (Randomized Condition) and a chosen subset of agents being aware of the memes (Hypothesis Condition). The Hypothesis Condition started with a fairly homogeneous subset of agents aware of each meme, who would be more prone to express the meme. The Randomized Condition started with a random set of agents aware of each meme, so there was less initial predisposition to spread the meme but might reach more people.

Table 7.38 shows some basic demographic information about the agents who know the meme in the Hypothesis condition, for reference. The GiveInformation initial agents are mainly Heremat members, whose group has a weak positive relationship with the US group. The Heremat group is well-positioned in society, but is a minority group with only 11 members. By comparison, the Shumar group has 38 members and the Yousif group has 23 members. The agents starting with GiveInformation have jobs such as public officials and policemen, or had experience in these fields before becoming unemployed. The PlantIED meme starts with a group of militants, some of whom are also tradesmen. Only half of them are currently employed and they strongly dislike the US Group.

Table 7.38: Hypothesis Condition: Meme-Aware Agent Demographics

Demographics	**GiveInformation**	**PlantIED**
% Shumar Group	16.3%	33.3%
% Heremat Group	66.7%	0%
% Yousif Group	16.3%	66.7%
% Employed	66.7%	50%
Avg. Valence Toward US Group	0.07	-0.6
Authority	0.17	0.0

Figure 7.16 shows the percentage of each group that learned the GiveInformation meme over time, as the mean of the 20 runs done in this condition. The x-axis shows the number of events that occurred within the simulation, which correlates with time passing. To avoid bias from the initial set of agents aware of the meme, this chart only considers agents who did not start knowing the meme. To help examine the learning region, this chart is truncated at the point where saturation was typically reached (all agents aware of the meme).

Next to it, Figure 7.17 shows this same statistic for the Randomized Condition.

While these charts only show the mean value across runs, the individual runs tended to be much more similar to their own mean run than to the mean runs of other groups. As such, the mean run seems to show a typical progression in this case. While the difference between each group's awareness of each meme is small, it was persistent across runs. Additionally, it should be stressed that these Figures include only new agents learning the meme and exclude the initial set. When taking into account the initial set, the differences between groups are far larger for the Hypothesis condition.

Comparing Figure 7.16 and Figure 7.17, it is evident that changing the initial set of agents changes the adoption curve of each group. Under the Hypothesis condition, GiveInformation is initially known by a significant number of Heremat agents. Due to this initial advantage, other Heremat agents tend to learn the meme more. In the Randomized Condition, this adoption advantage reverses. The Yousif group members and the Shumar group members tend to show advantages in learning the GiveInformation meme. In both conditions, the difference in learning only holds through the early adopter and early majority phases. Once the late majority phase starts, no particular group shows a significant advantage. Despite which group has an advantage, the diffusion rate of the meme is fairly similar- reaching saturation after approximately 1250 events (a bit more than a day).

Figure 7.16: % of Group Learned GiveInformation (Hypothesis Cond.)

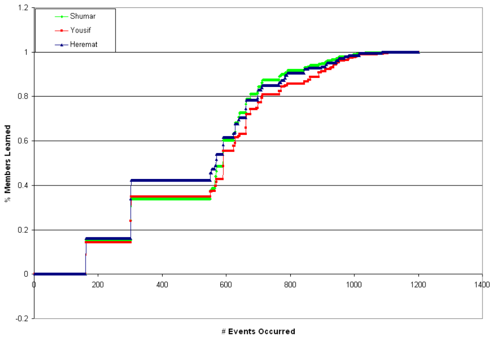
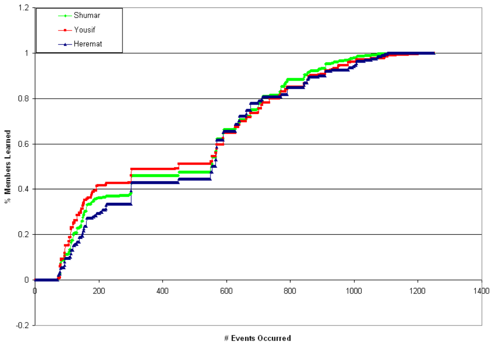


Figure 7.17: % of Group Learned GiveInformation (Randomized Condition)



The same comparison is shown for the PlantIED action, shown in Figure 7.18 (Hypothesis) and Figure 7.19 (Randomized). In both conditions, the Yousif group had an advantage in this meme. For the Hypothesis condition, a significant number of the initial carriers are members of the Yousif group. This allows them to better transmit the meme among their own group. However, even in the Randomized condition the Yousif group was slightly more favored in learning

Figure 7.18: % of Group Learned PlantIED (Hypothesis Condition)

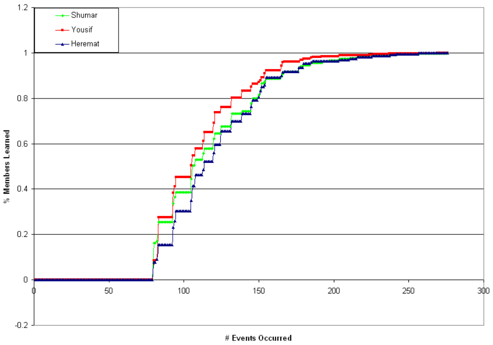
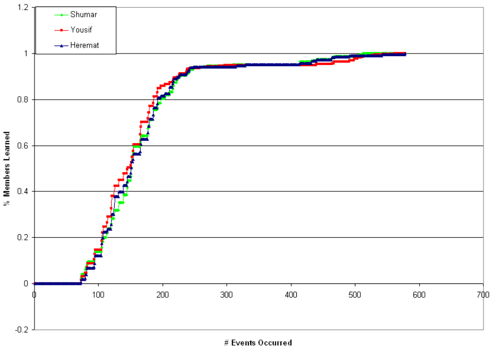


Figure 7.19: % of Group Learned PlantIED (Randomized Condition)

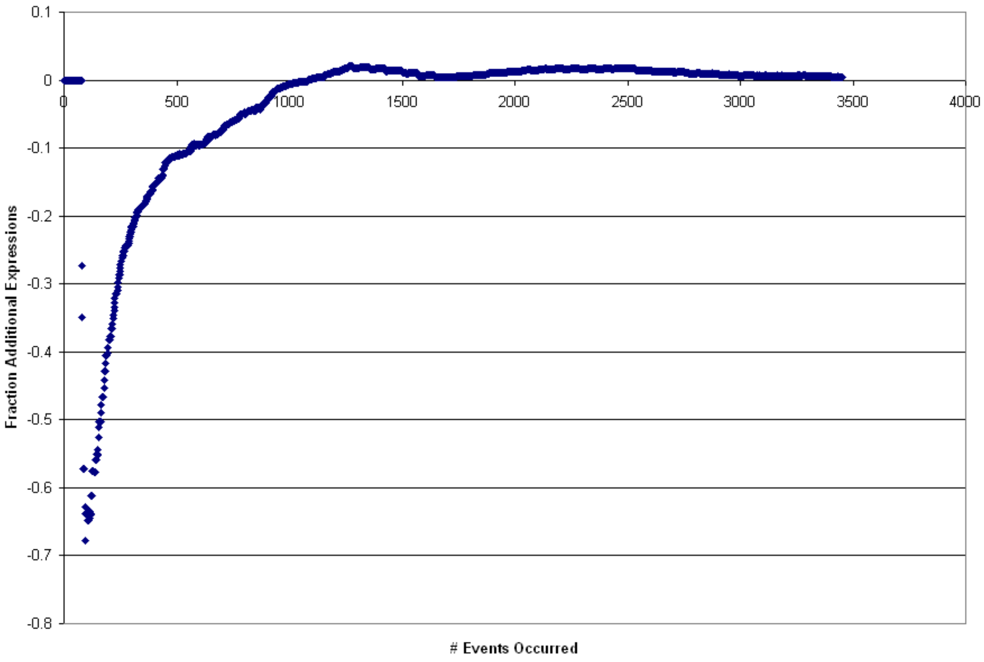


the meme. This indicates that the Yousif are in general more likely to learn this meme. Additionally, the rate of learning the PlantIED meme was greatly impacted by the starting condition. When given to a random set of agents, learning of this meme takes twice as long to saturate the population. It is also slower during the steeper learning curve, consistently lagging behind. This means that the starting set for PlantIED is more successful in getting awareness of that meme to the population than a random subset of agents would be.

$$RelativeExpression_H^R(t) = \frac{\sum_t \{\#Expressions_R(t) - \#Expressions_H(t)\}}{\sum_t \#Expressions_H(t)} \tag{7.13}$$

This effect could either be due to an increased effectiveness of the starting agents in spreading that meme, or it could indicate that those agents simply express the meme more frequently. To examine this, a comparison was made between number of expressions of PlantIED in the Hypothesis Condition as compared to the Randomized Condition. Figure 7.20 charts the fraction of difference in expressions between the Hypothesis Condition and the Randomized Condition, for the cumulative number of expressions up to that time point. This means that for each point at time t on this graph, the value is determined by Equation 7.13 where H represents the Hypothesis Condition, R represents the Randomized Condition, and $\#Expressions_X(t)$ represents the number of expressions that occurred during that time event period, on average, for a Condition X. Looking at this graph, it is clear that during the Hypothesis condition there are more expressions during the early portion of the simulation runs. These additional expressions are at least partly due to the Yousif group's strong negative relationship (valence) with the US Group. This could account for the advantage to learning conferred by having the Hypothesis agents aware of the meme.

Figure 7.20: Fraction of Difference of Cumulative PlantIED Expressions, (Randomized-Hypothesis)/Hypothesis



The alternative possibility for why PlantIED has much faster learning during the Hypothesis case is that other agents are more prone to learning from the starting set for this condition than they would be for a random set of agents. To look at this possibility, exposure efficiency graphs were generated. These graphs follow the same format as those in Section 7.4.2. Figure 7.21 and Figure 7.22 show the exposure efficiency graphs for the Hypothesis and Randomized conditions, respectively. Looking at these charts, the Hypothesis condition has a significantly higher efficiency for the first three or four agents who learn the meme. While this may be a modest gain, a small gain early in a diffusion process can lead to a faster tipping-point overall. From this analysis, PlantIED spreads better in the Hypothesis condition because the initial set of agents express it more and because other agents pay attention to them more when they express that meme.

Figures 7.21 and 7.22 also show slightly U-shaped curve as was noted in the Stanford diffusion analysis. It seems that this may be a general dynamic of how memes spread, based on the cognitive model design for the agents. To verify this, the exposure efficiency charts for GiveInformation are shown as Figure 7.23 and Figure 7.24 for the Hypothesis and Randomized conditions respectively. These charts also indicate a slight U-shaped curve. Unlike the PlantIED action,

Figure 7.21: Exposure Efficiency for PlantIED (Hypothesis)

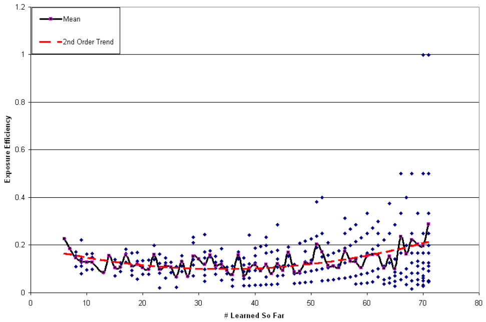
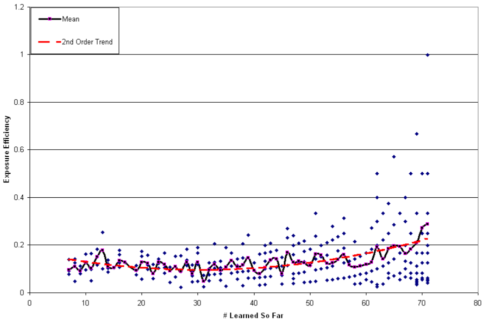


Figure 7.22: Exposure Efficiency for Plant IED (Randomized)



however, there is little advantage for GiveInformation during the Hypothesis condition in terms of exposure efficiency. This indicates that agents starting with GiveInformation in the Hypothesis condition were not as effective in spreading their meme. This is probably due to many of the initial members belonging to Heremat, a minority group with only 11 total members.

Figure 7.23: Exposure Efficiency for GiveInformation (Hypothesis)

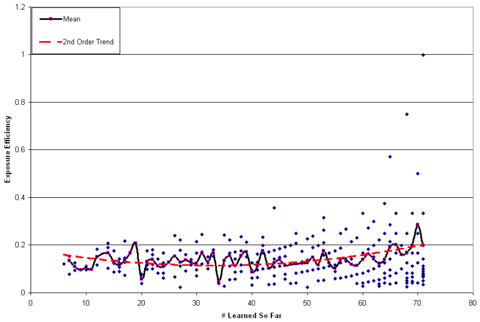
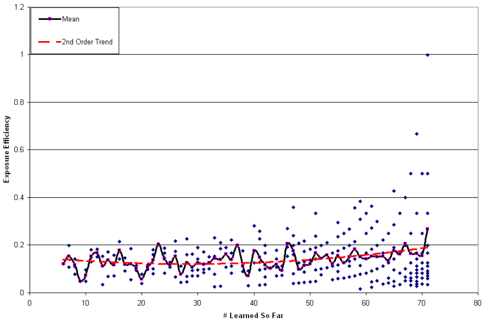


Figure 7.24: Exposure Efficiency for GiveInformation (Randomized)



This analysis indicates that diffusion of learning of these memes is occurring within the model. Comparing the two memes against each other, it appears that PlantIED transmits through the population much faster than GiveInformation. Looking at the relative number of expressions, however, PlantIED has only a modest advantage. The ratio of PlantIED to GiveInformation is 52:48 on average during the runs, with a standard deviation of approximately 0.8 for the ratio. A t-test was run to test for the probability that there were more PlantIED actions than GiveInformation actions for both experimental conditions, with the results shown in Table 7.39. The samples for this test were the number of expressions for each meme on each run, so the degrees of freedom for the test were 19 since there were 20 runs. The t-test strongly indicates that PlantIED was more common than GiveInformation. For reference, the table displays the confidence interval and average additional expressions of PlantIED compared to GiveInformation. This indicates that PlantIED is a significantly more popular option than GiveInformation in both conditions. It also indicates that the Hypothesis Condition has a slightly higher advantage, due to the initial set of agents as explained earlier.

Table 7.39: Expression Comparison of PlantIED vs GiveInformation

Metric	**Hypothesis**	**Randomized**
t-test: p-value	$7{\times}10^{-9}$	$7{\times}8^{-7}$
Avg(# PlantIED - GiveInformation)	31	27

Having established which meme was more successful within the village, the next question to answer is why. The next section will examine the factors that make each meme desirable to different groups.

7.5.2 Meme Transmission Dynamics

This section explores the who and why of meme transmission. The first step was figuring out which agents were most influential in spreading memes, a classification problem. From these classes, cross-group comparisons were performed to determine which personality and social properties differed significantly between the classes. This analysis only considered initial-condition factors, so these properties can be used as a priori profiles for key agents in the transmission of each meme.

In order to establish who are key agents in spreading each meme, agents were examined using the adoption factors presented during the Stanford Transmission Dynamics discussion (Section 7.4.2). These factors were: average time of first learning and average time of first expression. As before, first learning and first expression were examined together. Since this simulation is much larger than the Stanford simulation, quartiles were not used. Instead, the mClust clustering algorithm was used to generate an optimal set of clusters based upon the pair of variables. Unlike the previous section, the innovators (agents initially knowing the meme) are included in this analysis.

To examine the differences between these clusters, a set of demographic properties was collected from the agents belonging to each cluster. The set of properties used for clustering are shown in Table 7.40. These properties include GSP personality tree factors, group memberships, valences toward other groups, authority, and employment level. The full set of GSP nodes is not enumerated in this section for the sake of brevity, but a brief summary of each node is contained

Table 7.40: Demographic Properties for Cross-Cluster Analysis

Property	**Data Type**	**Description**
Group Valences		
- Valence(US Group)	Continuous	Like/dislike toward the US group
- Valence(Heremat Group)	Continuous	Like/dislike toward the Heremat group
- Valence(Shumar Group)	Continuous	Like/dislike toward the Yousif group
- Valence(Yousif Group)	Continuous	Like/dislike toward the Yousif group
Group Memberships		
- Member of Heremat	Dichotomous	True only if agent in Heremat faction
- Member of Shumar	Dichotomous	True only if agent in Shumar faction
- Member of Yousif	Dichotomous	True only if agent in Yousif faction
Social Properties		
- Authority	Continuous	Authority of the agent in his/her group
- EmploymentLevel	Dichotomous	If True, agent is employed and typically goes to work during the day
GSP Personality Factors	Continuous	Personality traits, as defined in Appendix H

in Appendix H, Table H.1. Instead, each node will be briefly described in-text if it shows a particular significance for analysis. For all continuous properties, a one-way ANOVA was run to detect any significant differences between clusters. After this, a Scheffe post-hoc test was applied in order to examine the specific differences between individual clusters. For dichotomous variables, a chi-squared test was run to detect differences.

7.5.3 Adoption Indicators

This section focuses on adoption: the first learning and expression of each meme. Since there is no assurance that an agent will learn or express the meme, an agent who never expresses the meme is placed at the last step of the simulation for the purposes of averaging (step 3456). Any cluster in which no members expressed the meme will be labeled as “Never” to differentiate it from clusters in which some members expressed the meme. The clustering results for GiveInformation in the Hypothesis and Randomized Conditions are shown in Figure 7.25 and Figure 7.26 respectively. The diffusion between these conditions is different not only in the structure of the clusters, but in the number of clusters overall. The Hypothesis condition shows a total of 5 clusters. Clusters will be referred to by their means during the discussion, in the form (First Learning Time, First Expression Time).

The cluster in the lower left hand (0,256) is the initial set of agents aware of the meme, who tend to express it relatively early. At the upper right hand of the graph (517,3448) is a significant number of agents who learn the GiveInformation meme late and most never express it. Of the remaining three clusters, the those centered at (517,993) and (487,2580) were diffuse but (581,1284) was very dense. The Randomized condition was much simpler- containing only two diffuse groups for learning and expression located at (412,956) and (419,2983). Interestingly in this case, both clusters have similar learning time centers but very different expression times.

Figure 7.25: GiveInformation First Learning and First Expression Clusters (Hypothesis Condition)

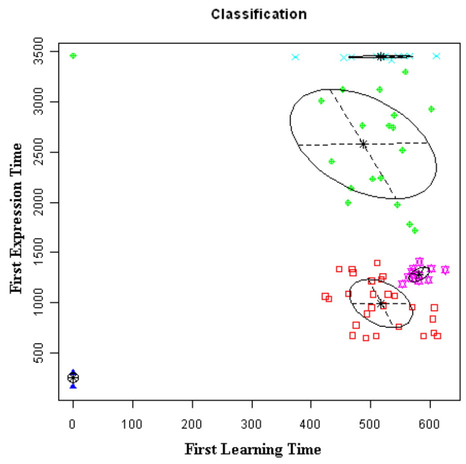


Figure 7.26: GiveInformation First Learning and First Expression Clusters (Randomized Condition)

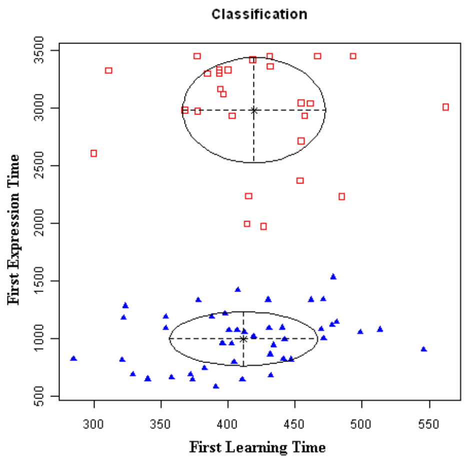


Table 7.41: Demographics for GiveInformation Learning and First Expression Clusters (Hypothesis Condition)

Cluster At	Cluster Size	Primary Groups	Learning Adoption	Expression Adoption
(0,256)	3	Heremat	Innovator	Early Adopter
(517,994)	27	Shumar	Late Majority	Early Majority
(581,1284)	12	Yousif	Laggard	Late Majority
(487,2580)	19	Shumar, Heremat	Early Majority	Laggard
(517,3448)	11	Yousif, Shumar	Late Majority	Holdout

Table 7.41 shows the basic information about each cluster, including its size and dominant groups represented. Also, each cluster is categorized into its adoption category. One additional category is used here that was not used in the Stanford analysis, which is the Holdout category. These agents generally did not express the meme at all. In this respect, they were not laggards but simply were unlikely to express GiveInformation at all.

GiveInformation first learning and expression time clusters were analyzed using an ANOVA based on five groups, one for each cluster, for all continuous properties noted earlier in Table 7.40. Discrete properties noted in that same table were each analyzed using chi-squared tests. A very large number of differences were statistically significant between clusters ($p < 0.05$), so only those that most uniquely identified each cluster will be discussed. As such, key identifier properties noted are significant at the 0.05 level in differentiating them from other clusters, based upon the Scheffe post-hoc test. The largest deviations between clusters were shown for the two corner clusters: (0,256) and (517,3448). In addition to being different from each other, they were both significantly different from other groups on a number of measures. By comparison, the clusters centered at (412,956), (419,2983), and (581,1284) were fairly similar. Table 7.42 shows the properties which distinguish groups from other groups in the scenario.

Table 7.42: Key Identifiers for GiveInformation Learning and First Expression Clusters (Hypothesis Condition)

Cluster At	**Property**	**Defining Characteristics**
(0,256)	Valence(User Group)	Likes User Group more than others (0.067 vs -0.47 outside cluster)
	Group Membership	Primarily Heremat (2 out of 3)
	GSP(Be_Task_Focused)	Less focused on problem solving, more focused on building relationships
	GSP(Physiology)	Values basic needs more than others.
	GSP(None_r_Sensitive)	Less regard for human life than others
(517,994)	Group Membership	Dominantly Shumar (75% of cluster).
	GSP(Friendly_Faction)	Less happy to see friendly factions succeed
(581,1284)	Group Membership	Dominantly Yousif (75% of cluster)
	GSP(Symbolic)	Values symbolic payoffs less
(487, 2580)	Group Membership	More Heremat than other groups (5 members out of 19)
	Valence(Yousif)	Dislikes the Yousif group, especially compared to (581, 1284)
(517,3448)	GSP(Safety)	Values personal safety more
	GSP(For_The_Self)	High preference to preserve self
	GSP(Life_Res_r_Sensitive)	High value for human life
	GSP(Help...)	Higher value for all “Help” GSP nodes, regardless of group
	GSP(Enemy_Is_Outgroup)	Less inclined to treating enemies poorly
	GSP(Esteem)	Less value for self-efficacy
	GSP(Be_Controlling)	Values sense of control much less
	GSP(Symbolic)	Values symbolic payoffs more

Looking at the clusters, it appears that group membership and personality are the strongest determinants of being in a particular cluster. The innovator cluster at (0,256) is small and not very influential. Even among agents initially aware of the GiveInformation meme, not all of them reliably express it. The reliably different characteristic of this subgroup is that it likes the US Group. All other clusters dislike the US, to varying degrees. There are some other significant personality differences, but these may be unique to the small sample size for that cluster. Most of the remaining Heremat members are part of the (487, 2580) cluster. These agents are some of the first ones to learn the meme but among the last to try it. One of the differences between the innovator group and this cluster is that the innovators give a higher importance to relationship-building. A more task-focused agent will tend to see less value in contacting a third party to express problems.

At the opposite end of the spectrum is the holdout cluster at (517,3448). This cluster is not very different from the other three more moderate clusters in group membership or in learning time. Examining the differences from the ANOVA analysis, it is clear why this group does not express the GiveInformation meme. Members of this group place a very high value on personal safety and preferences for the self (high Safety goals and For_the_Self preferences). They also have a much lower inclination to control their environment, as shown by low importances for Esteem and Be_Controlling. Overall, this cluster of agents shares a personality type that is not inclined to take risks. Considering that becoming an informant is a dangerous endeavor, these agents would simply rather stay home.

The remaining clusters for the Hypothesis Condition are gathered tightly around group membership. (517,994) is a Shumar-dominated group and (581,1284) is a Yousif-dominated group. These clusters are not as well distinguished from the other clusters, both in the cluster sharpness and in demographic properties. The difference in learning time is probably explained by the poor relationship between the Yousif group and Heremat group, who dominate the innovators. In general, this analysis does not capture the factors that appear causal to these smaller differences, however. It seems likely that these clusters form due to a mixture of factors, rather than the strong indicators seen for the other groups.

Table 7.43: Demographics for GiveInformation Learning and First Expression Clusters (Randomized Condition)

Cluster At	Cluster Size	Primary Groups	Learning Adoption	Expression Adoption
(412,996)	45	Mixed	Majority	Early Majority
(419,2983)	27	Mixed	Majority	Late Majority

By contrast, the Randomized Condition shows a much flatter and more diffuse set of clusters. Table 7.43 shows the basic information for the clustering of the Randomized Condition shown in Figure 7.26. The differences between these clusters lie almost entirely on the time of first expression. The earlier group expresses during the first third of the simulation, while the other group consists of agents who express much later or not at all.

The trends seen in the Hypothesis condition disappear in this condition. Without the initial social biases, the memes spread across groups fairly evenly. Between these clusters, the only significant differences were in personality. Group membership and even group valence toward the US were not as significant as the personality factors leading agents to be willing to provide information to the US. The most influential factors for determining membership in the earlier expressing cluster high levels of the GSP personality traits: Be_controlling, Be_Open, Bring_About_Greater_Good, Assert_Individuality, and Physiology. This indicates a personality type that seems prepared to bring about changes and is concerned with matters of power and control. While Be_Open and Be_Controlling are technically competing nodes on the GSP tree, they come from a common node about how to exercise power and control. As such, it seems reasonable that both could be positive expressing this meme. The high level for The high importance of Physiology goals appears to be due to a negative correlation with Safety goals. Since a low importance of personal Safety is one of the strongest indicators for early meme expression, this appears to be a secondary indicator.

On the converse side, traits that most associated with late expression were: Safety, Materialism, Respect for Life, Keep One's Word, and Grow Economy. This reinforces the findings from the Hypothesis condition, which indicates that agents strongly concerned with safety and material goods will tend to avoid giving information if possible.

The PlantIED meme showed some similarities in its learning and first expression dynamics. Figure 7.27 and Figure 7.28 show the mClust cluster graphs for PlantIED for the Hypothesis and Randomized conditions, respectively. As with the GiveInformation meme, the Hypothesis Condition showed much cleaner clusters than the Randomized Condition. However, the Randomized Condition for PlantIED showed much more nuanced behavior than the Randomized Condition for GiveInformation.

Table 7.44 shows the basic demographics for the Hypothesis clusters and their approximate adoption positions. Even more so than GiveInformation in the Hypothesis condition, the clusters closely correlate with group membership. The majority of Shumar and Heremat learn the meme later and wait much longer to express it, if at all. Conversely, a subset of the Shumar and Yousif quickly move to express the meme. The PlantIED meme in this Condition is interesting because learning and first expression track each other quite closely. The agents who are last to learn this meme are also the least likely to want to express it. This

Figure 7.27: PlantIED First Learning and First Expression Clusters (Hypothesis Condition)

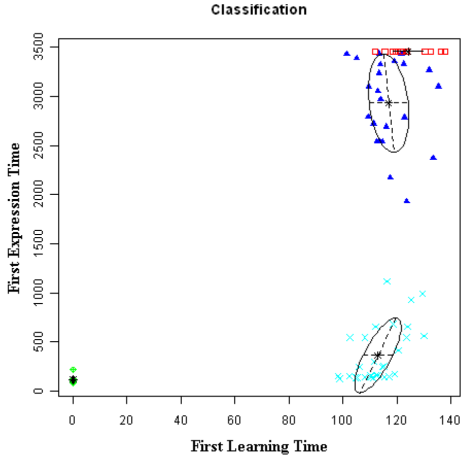


Figure 7.28: PlantIED First Learning and First Expression Clusters (Randomized Condition)

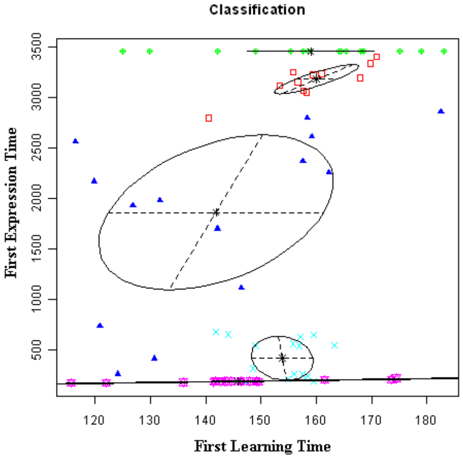


Table 7.44: Demographics for PlantIED Learning and First Expression Clusters (Hypothesis Condition)

Cluster At	Cluster Size	Primary Groups	Learning Adoption	Expression Adoption
(0,116)	6	Yousif, Shumar	Innovator	Early Adopter
(113,361)	30	Shumar, Yousif	Early Adopter	Early Majority
(117,2931)	23	Shumar, Heremat	Early Majority	Late Majority
(124,Never)	13	Shumar, Heremat	Late Majority	Holdout

is at a contrast with GiveInformation, where holdouts uninterested in expressing the meme still learned it about as fast as other agents. In this case, attention correlates well with the motivation to imitate.

Table 7.45 shows the identifiers for the clusters for the PlantIED action under the Hypothesis condition. This analysis shows a fairly sharp distinction between the types of agents in each of these categories. The (0,116) innovators of the PlantIED action are prone to expressing the meme because they feel it will benefit their group's future, as well as to satisfy their own needs for esteem and asserting their individuality. They also place a low importance on their own safety. They are also primarily Yousif group members, and share a negative valence toward the US Group.

The cluster at (113,361) is similar, with low importance weights on safety and heightened weights on individuality and esteem. However, they differ slightly in that their long term preferences are more oriented toward symbolic outcomes

Table 7.45: Key Indicators for Learning and First Expression PlantIED Clusters (Hypothesis Condition)

Cluster At	Property	Defining Characteristics
(0,116)	GSP(Safety)	Personal safety is much less important
	GSP(Esteem)	Respect and esteem is very important
	GSP(Assert_Individuality)	Individuality more important
	GSP(For_the_Group)	Good long-term future for group more important
	GSP(Use_Asymmetric_Attacks)	More prone to using asymmetric maneuvers
	GSP(Belonging)	Higher need for belonging
(113,361)	GSP(Assert_Individuality)	Individuality more important
	GSP(Esteem)	Respect and esteem more important
	GSP(Symbolistic)	Symbolic outcomes more important
	GSP(Use_Asymmetric_Attacks)	More prone to using asymmetric maneuvers
(117,2931)	Valence(User Group)*	More positive toward US group (*not Scheffe-significant)
	GSP(GG_Economy)	Greater importance to economic growth
	GSP(For_the_Self)	More interested in a good future for the self
	GSP(For_Everybody)	More interested in everyone's future
	GSP(Conform_to_Society)	More prone to conforming
	GSP(Own_People)	Less interested in ingroup's future
(124,Never)	GSP(Safety)	Very concerned with personal safety
	GSP(Life_Res_r_Sensitive)	Strong respect for life
	GSP(For_the_Self)	More interested in good future for self
	GSP(Respect_Authority)	Higher respect for authority
	GSP(Be_Task_Focused)	Lower task focus, higher relationship focus.
	GSP(Bring_About_Greater_Good)	More interested in good for all
	GSP(Enemy_is_Out_Group)	Less inclined to discriminate against enemies
	GSP(Materialistic)	Less interested in material goods

rather than outcomes that benefit their own group. This difference appears to be influence of the Al Qaeda Iraqi (AQI) members in the village, from looking at the individual agents. Overall, these agents waste little time between learning of the meme before volunteering to plant an IED.

The clusters that resist the meme are quite different in nature. The (117,2931) cluster, which is partially resistant to expressing the meme, appears to be business-oriented. It places high importance on growing economic resources, on conforming to society, and on getting positive outcomes for the self. It also places a higher importance on safety than the IED-active clusters, but not as high as the other resistant cluster at (124,Never). Additionally this cluster has the highest opinion of the US group, the only cluster that conclusively likes the US Group.

The cluster at (124,Never) contains resistant agents who appear to be self-interested good guys. Their primary identifying characteristics are that they are interested in their own safety and the respect of life in general. They are also much more interested in their personal long-term outcomes than the PlantIED adopters. However, they also tend to be less materialistic, more interested in the greater good, and assign less value to treating outgroups badly. Interestingly, this cluster does not include any of the GiveInformation innovators nor does it disproportionately favor agents that tended to express GiveInformation significantly earlier. As such, some of these members actually overlap with the resistant agents for the GiveInformation meme. This is probably because both memes both have negative activations for safety.

Table 7.46 shows the amount of overlap between each cluster from the GiveInformation Hypothesis condition and each cluster in the PlantIED Hypothesis condition. This supports the prior intuition that many of the same agents tend to not express in both cases. Additionally, this table also shows that late expression agents did not tend to become early expression agents, for either case. It also shows that a significant number of agents willing to plant IEDs might also be willing to give information tips to the US, in certain circumstances (e.g., when gangs inform on each other).

Table 7.46: Hypothesis Cluster Overlap for Learning and First Expression Times, GiveInformation vs. PlantIED

			GiveInformation		
PlantIED	(0,256)	(517,994)	(581,1284)	(487,2580)	(517,3448)
(0,116)	0	2	3	1	0
(113,361)	0	12	9	9	0
(117,2931)	3	13	0	5	2
(124,Never)	0	0	0	4	9

The Randomized Condition for PlantIED also shows interesting behavior. Table 7.47 shows the cluster overlap for the clusters in the Randomized Condition with those in the Hypothesis condition of PlantIED. This table shows that the Randomized Condition does significantly influence the learning and first expression times. As with the GiveInformation action, using Randomized initial sets washes out most of the differences in learning- leaving only minor differences between the clusters. For both memes, the Hypothesis Condition shows a strong bias in learning where agents in the same group tend to learn together. The internal validity analysis performed much earlier as part of Section 7.3.2 showed that this interaction also existed in the Randomized Condition, where attention correlated with membership in the same ingroup. However since a different set of agents start with the meme for each run in the Randomized Condition, no group had an advantage for learning. This shows that no group is innately more likely to learn each meme, but that the biases in learning result from memes being transmitted through the social structure.

Some agents have also moved from typically expressing earlier to typically expressing later. For many of the cases, this is a small re-shuffling, but in other cases it may lead to more significant consequences. For example, the Hypothesis cluster at (117,2931) breaks into two smaller clusters. One of those clusters (160,3183) has a much longer amount of time before first expression while the other includes four additional subjects (142,1861). Table 7.48 shows the basic demographics for the PlantIED action under the Randomized Condition, for comparison against Table 7.45 which contained the Hypothesis data.

Table 7.47: PlantIED Cluster Overlap for Learning and First Expression Times

		Randomized			
Hypothesis	(146,190)	(154,417)	(142,1861)	(160,3183)	(159,Never)
(0,116)	4	2	0	0	0
(113,361)	13	13	4	0	0
(117,2931)	3	0	10	11	2
(124,Never)	0	0	0	0	13

Table 7.48: Demographics for PlantIED Learning and First Expression Clusters (Randomized Condition)

Cluster At	**Cluster Size**	**Primary Groups**	**Learning Adoption**	**Expression Adoption**
(146,190)	17	Yousif, Shumar	Early Majority	Early Adopter
(154,417)	15	Shumar, Yousif	Late Majority	Early Majority
(142,1861)	14	Shumar, Heremat	Early Majority	Late Majority
(160,3183)	11	Mixed	Late Majority	Laggard
(159,Never)	15	Shumar, Heremat	Late Majority	Holdout

Looking at the ANOVA and chi-squared analysis, it appears that the Randomized Condition leads to shifts in the cluster indicators as well. For example, the cluster at (142, 1861) has higher EmploymentLevel and Authority level compared to other groups. These additional work responsibilities may play a role in that subgroup's delay in first expression. Most of the prior indicators of early or late first expression still hold. In particular, higher personality traits for Assert_Individuality Belonging, Esteem, Enemy_is_Outgroup, and For_the_Group are still solid indicators that an agent may be more likely to express the PlantIED meme. Conversely, For_the_Self and Safety are still good indicators that an agent will not tend to express the PlantIED meme. While these indicators get stronger, weaker indicators such as Conform_to_Society wash out. Those indicators were probably unique to that particular configuration and are not reliable predictors for the meme in general. Alternatively, this might mean that those indicators were related to learning rather than expression. The next section summarizes the indicators which were reliable for both the Hypothesis Condition and Randomized Condition, for learning and first expression times. These will be referred to as the Key Indicators for the type of agent and situation which leads agents to adopt GiveInformation or PlantIED.

7.5.4 Key Indicators for Meme Adoption

The prior analysis showed that it is possible to determine the statistically significant differences between faster versus slower adopters, on different metrics. Table 7.49 summarizes the key indicators that differentiated early learners versus late learners, for the GiveInformation and PlantIED memes. From this analysis, the early learners tend to be differentiated primarily by their social network (e.g., ingroup) which accounts for most of the variance in learning. Their interest in the meme is a small secondary influence on top of this. The Randomized Condition cases showed that giving a meme to a purely random subset tends to lead to fairly equal learning rates, on average.

Table 7.49: Key Indicators for Determining Meme Learning in Iraqi Village

Key Indicator	**Give Information Learning Time Change**	**PlantIED Learning Time Change**
Same Ingroup as Innovators	Faster Learning	Faster Learning
Group Likes Innovator Group	Slightly Faster Learning	Slightly Faster Learning
Innovators Express Earlier	Faster Learning	Faster Learning
Less Prone to Express Meme	No Clear Connection	Slightly Slower Learning

On the converse, personality factors dominate which agents tend to express memes earlier. Table 7.50 shows the key indicators that help determine if an agent will express a meme earlier or later. Membership in a group which likes or dislikes the US Group is the only consistent non-personality key factor that influences expression of either meme in this simulation. Higher employment may have also been an environmental influence, but was not statistically significant. Otherwise, expression was almost entirely determined by the personality factors. Safety goals were a key limiting factor for both memes, an obvious connection for dangerous actions. However, seemingly unrelated factors such as long term preferences for oneself and materialism have a significant influence as well. This indicates that these memes are competing with day to day activities such as going to work and pursuing economic endeavors.

The inferred factors that affect learning and expression of these memes appear to form a reasonable set of properties within this situation which could. While

Table 7.50: Key Indicators for Determining Meme First Expression in Iraqi Village

Key Indicator	**Give Information First Expression Time**	**PlantIED First Expression Time**
↑ Valence Toward US	Earlier expression	Slower expression (or None)
GSP Goals (Short Term Values)		
↑ Safety	Prevents expression	Prevents expression
↑ Esteem	-	Earlier expression
GSP Standards (Preferred Methods)		
↑ Assert Individuality	Earlier expression	Earlier expression
↑ Be Task Focused	Slower expression	Earlier expression
↑ Be Controlling	Earlier expression	Earlier expression
↑ Bring About Greater Good	Earlier Expression	-
↑ Use Asymmetric Attacks	-	Earlier expression
GSP Preferences (Long Term Wants)		
↑ For Own Group	-	Earlier Expression
↑ For the Self	Slower expression	Slower expression
↑ Materialistic	Slower expression	Slower expression

the Stanford Experiment showed that a real-world situation could be modeled with reasonable fidelity, this simulation has shown that this model allows the study of meme competition. Analysis of the factors that promote certain memes in an environment gives insight into the personality and societal conditions that promote or stifle certain memes. If the model can be shown to have predictive value for modeling meme selection based upon real-world data, this could be a powerful tool for examining trends that emerge in different sub-populations.

Chapter 8

Conclusions

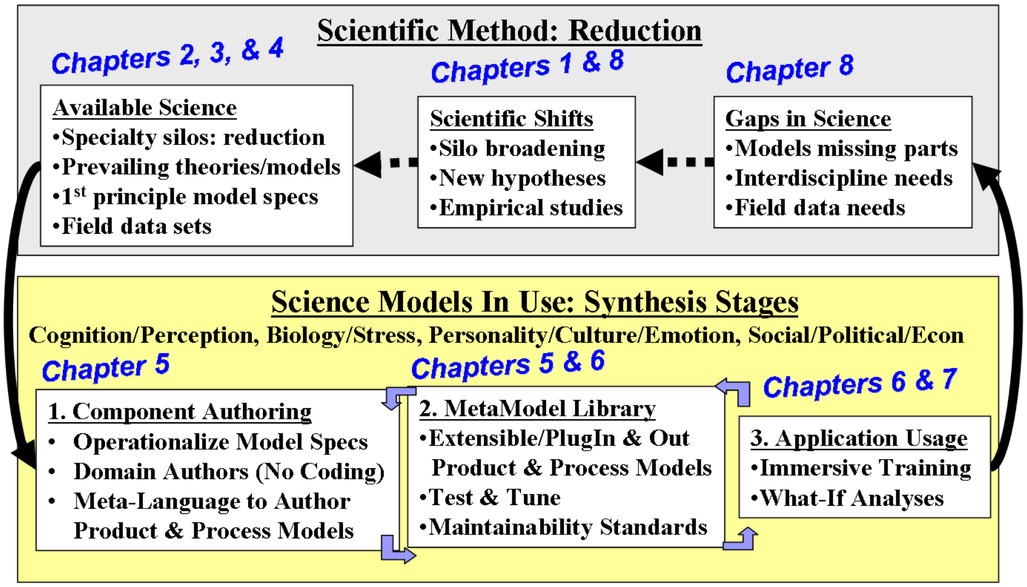
This thesis set out to make a contribution to modeling memes, units of cultural meaning that reproduce recursively through society. Memetics can be an effective framework for studying the evolution of ideas and culturally transmitted practices. However, memetics is still a relatively young approach to studying culture and has been missing solid formalisms for the definition of memes, as well as a solid connection to the cognitive factors that influence social transmission of information. The main goal of this work was to expand the study of memes to include relevant empirical research explaining the environmental and cognitive mechanisms that drive meme evolution. The approach taken was to bring a systems perspective to memes in order to develop a useful architecture for modeling memes. This architecture was intended to be complete, holistic, and workable for studying meme transmission and selection pressure.

8.1 Overview of Contributions

As stated in Chapter 1, the approach taken by this work was based on a Systems Social Science development cycle. For reference, Figure 8.1, a duplicate of Figure 1.2, shows the development cycle used by this research. Each of the stages of this development cycle was completed, helping to move from narrow bands of focused knowledge toward a model capable of representing memes in a social system.

Each of these steps has provided meaningful contributions to the study of memes. In the process, this research has also provided insight into general social science questions and has produced a novel analytical technique. Figure 8.2 notes the major contributions of this thesis, categorized by their stage in the systems social science inquiry approach. These contributions will each be discussed briefly, examining the significance of the contribution and how it might assist further scientific endeavors.

Figure 8.1: Systems Social Science Development Cycle. Adapted from Silverman (2006)



Scientific Method: Reduction

Chapters 2, 3, & 4

Available Science
- Specialty silos: reduction
- Prevailing theories/models
- 1st principle model specs
- Field data sets

Chapters 1 & 8

Scientific Shifts
- Silo broadening
- New hypotheses
- Empirical studies

Chapter 8

Gaps in Science
- Models missing parts
- Interdiscipline needs
- Field data needs

Science Models In Use: Synthesis Stages

Cognition/Perception, Biology/Stress, Personality/Culture/Emotion, Social/Political/Econ

Chapter 5

1. Component Authoring
- Operationalize Model Specs
- Domain Authors (No Coding)
- Meta-Language to Author Product & Process Models

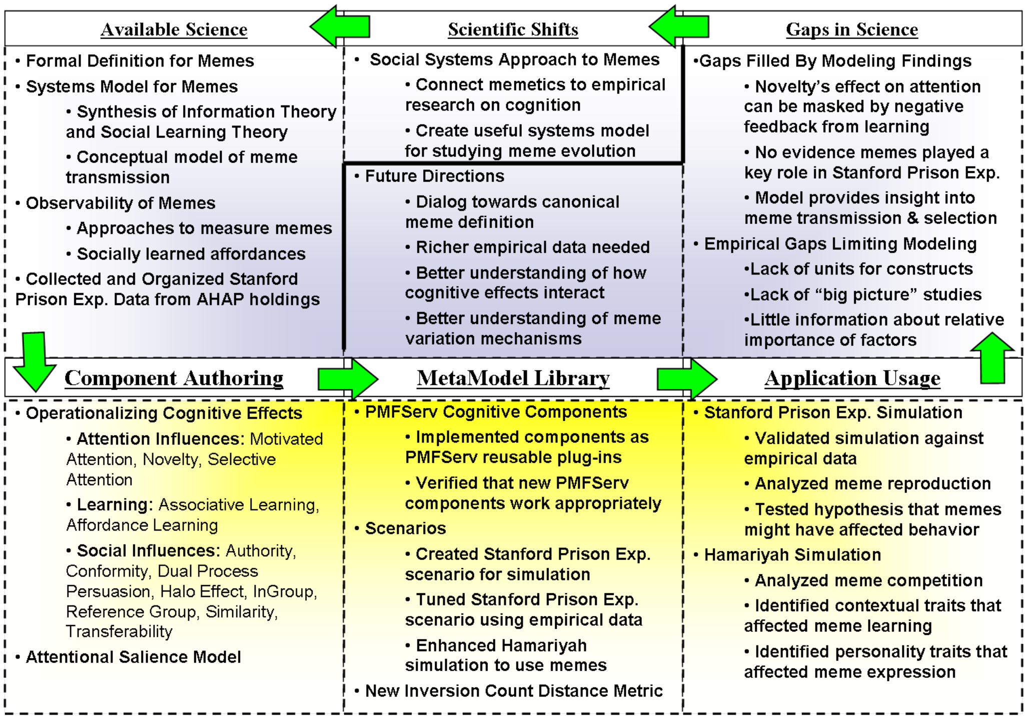
Chapters 5 & 6

2. MetaModel Library
- Extensible/PlugIn & Out Product & Process Models
- Test & Tune
- Maintainability Standards

Chapters 6 & 7

3. Application Usage
- Immersive Training
- What-If Analyses

Figure 8.2: Systems Social Science Development Contributions



Available Science
- Formal Definition for Memes
- Systems Model for Memes
  - Synthesis of Information Theory and Social Learning Theory
  - Conceptual model of meme transmission
- Observability of Memes
  - Approaches to measure memes
  - Socially learned affordances
- Collected and Organized Stanford Prison Exp. Data from AHAP holdings

Scientific Shifts
- Social Systems Approach to Memes
  - Connect memetics to empirical research on cognition
  - Create useful systems model for studying meme evolution
- Future Directions
  - Dialog towards canonical meme definition
  - Richer empirical data needed
  - Better understanding of how cognitive effects interact
  - Better understanding of meme variation mechanisms

Gaps in Science
- Gaps Filled By Modeling Findings
  - Novelty's effect on attention can be masked by negative feedback from learning
  - No evidence memes played a key role in Stanford Prison Exp.
  - Model provides insight into meme transmission & selection
- Empirical Gaps Limiting Modeling
  - Lack of units for constructs
  - Lack of "big picture" studies
  - Little information about relative importance of factors

Component Authoring
- Operationalizing Cognitive Effects
  - Attention Influences: Motivated Attention, Novelty, Selective Attention
  - Learning: Associative Learning, Affordance Learning
  - Social Influences: Authority, Conformity, Dual Process Persuasion, Halo Effect, InGroup, Reference Group, Similarity, Transferability
- Attentional Salience Model

MetaModel Library
- PMFServ Cognitive Components
  - Implemented components as PMFServ reusable plug-ins
  - Verified that new PMFServ components work appropriately
- Scenarios
  - Created Stanford Prison Exp. scenario for simulation
  - Tuned Stanford Prison Exp. scenario using empirical data
  - Enhanced Hamariyah simulation to use memes
- New Inversion Count Distance Metric

Application Usage
- Stanford Prison Exp. Simulation
  - Validated simulation against empirical data
  - Analyzed meme reproduction
  - Tested hypothesis that memes might have affected behavior
- Hamariyah Simulation
  - Analyzed meme competition
  - Identified contextual traits that affected meme learning
  - Identified personality traits that affected meme expression

8.2 Studying Available Science

By working with the available science, empirical and theoretical work was integrated to produce a few major contributions for the study of memes. This body of work is described in Chapters 2, 3, and 4. These three chapters consolidated and integrated a significant body of social science theories and specialized reductive findings, showing how available science could enrich the study of memes.

Firstly, a formal definition for memes was proposed which incorporated theoretical and pragmatic concerns about the meaning of a meme. The next major contribution was to synthesize a systems model for memes that connected Social Cognitive Learning Theory and Information Theory in order to model meme evolution (Bandura, 1986; Shannon, 1948). This contribution produced a conceptual model for meme transfer that integrated many of the cognitive components that explain meme transmission and selection. Additionally, this definition and model were used to gain insight into how memes might be measured and studied empirically. This is an important point for the viability of these contributions, since memes must be linked to observable phenomena in order to study them appropriately.

Finally, available science was also harnessed in order to model the Stanford Prison Experiment. This contribution involved collecting and organizing archival data from the original Stanford Prison Experiment. Through this work, a significant amount of information previously only recorded on paper records was digitized, de-identified, and organized. This work may assist future researchers seeking to examine with these holdings at the Archives of the History of American Psychology (AHAP) or seeking an additional perspective into the structure of the Stanford Prison Experiment.

8.2.1 Formal Definition for Memes

Chapter 2 set out a formal definition for memes, with consideration of both theoretical concerns and pragmatic concerns related to the empirical study of memes. A number of scholars have voiced the need for clear and testable definitions for memes (Blackmore, 1999; Finkelstein, 2008). In this literature review, no formalized definition of memes was found that was expressed algorithmically and was disprovable. The formalization put forth in this thesis addresses this need. This definition frames memes as a form of semantic information that is transmitted as the result of social behavior such as performing actions, verbal communication, or written signs. In this way, the definition is consistent with semiotic work on signs but also adds the additional constraint that a meme must be able to recursively reproduce within a population. Additionally, this definition is novel in that it defines a meme only in reference to a population.

Surprisingly, this has not been a major characteristic in defining memes despite the fact that different organisms and even age groups can differ significantly in how they process information.

This formalization for memes provided a good working definition for this thesis to build upon. Formal definitions are essential for examining memes, which could otherwise be fuzzy and not well specified. The definition presented attempted to balance issues of observability, the ability to measure memes, with ontological adequacy, the ability of the definition to include all possible memes. The definition created appears to adequately cover the concepts typically considered memes, making it ontologically adequate. It also lends itself to measurement, supporting observability. However, this definition does not necessarily imply that all possible memes are observable. Overall, this definition was extremely useful for framing analysis of memes cognitively and for connecting memes to the empirical domain.

Hopefully, this formal definition will be of use to other researchers attempting to model or measure memes. Additionally, it contributes to the larger ongoing discussion of how to define memes. This discussion is a difficult one because it requires balancing the philosophical concerns of ontological adequacy against the pragmatic concerns of scientific study and measurement. This is still a very active area. For example, Finkelstein (2008) recently proposed a pragmatic definition of memes in terms of their measurable qualities: how much they reproduce (propagation) and how long they reproduce (persistence). In this conference paper, a stated goal of his definition was to encourage dialog that would "eventually converge to a canonical definition that will be useful in establishing a scientific basis for memetics" Finkelstein (2008, p. 16). The definition from Chapter 2 provides a new and distinct viewpoint that can contribute significantly to a canonical definition. As such, this was an important part of the work within this thesis.

8.2.2 Systems Model for Memes

Chapter 3 used this definition as the basis for a systems model for examining meme evolution within a social system, drawing from information theory and social cognitive learning theory (Bandura, 1986; Shannon, 1948). A core intuition that has guided this work is that a synthesis between Bandura (1986) social learning theory and the Shannon (1948) information theory provides an effective framework for studying memes. Bandura's work, especially on observational learning, provides an effective framework for examining how social learning and imitation work at the cognitive level. Information theory provides an effective framework for examining how memes transmit through the environment as information. As shown in Chapter 3, these models can be connected to form a socio-cognitive environment for memes.

This systems model directly addresses the quote referenced in Chapter 3 by Castelfranchi (2001), “Memetics needs cognitive modeling.” From front to back, this approach to examining memes has been rooted in the domains of cognition, information processing, and cognitive modeling. This research identified a robust set of cognitive factors that influence meme reproduction, selection, and variation. In addition to synthesizing a systems model for examining memes, the literature review presented in Chapter 3 should be a useful resource for other researchers attempting to examine memes from a cognitive standpoint.

A major contribution of this research was to develop a useful architecture for studying memes. From a systems standpoint, usefulness requires three primary criteria: completeness, holism, and workability. Completeness requires that the system for examining memes must be able to explain each of its mechanisms. For the architecture to be complete, it must be able to represent evolution of memes. Holism requires the system to be more than the sum of its parts and that the system cannot be broken apart without losing essential interactions. For the architecture to be holistic, the constituent theories (Observational Learning and Information Theory) must meaningfully interact to provide insight beyond either one individually. Workability requires that the system be implementable and practical for approaching its intended purpose. For the system to be workable, it must be implementable and able to be applied to realistic problems.

This research has attained each of these goals and thus presents a useful architecture for modeling memes. The systems model for memes has been shown to be complete, holistic, and workable. The support for each of these claims is shown below.

Demonstration of Completeness

Memes work according to evolutionary mechanisms, so a complete architecture for modeling memes must be able to explain reproduction, variation, and selection pressure. Chapter 2 addresses each of these aspects. In particular, Sections 3.2, 3.3, and 3.4 present mechanisms within the system that enable reproduction, variation, and selection of memes.

Meme reproduction was explained as a natural outcome of the process of learning from the environment and producing behavior based upon social learning. In this context, the Bandura (1986) social learning theory explains how an agent learns memes from the environment and reproduces them back into the environment. Information theory provides an understanding of the processes that influence a meme while within the environment, such as the transmission medium and encoding (Shannon, 1948). Together, these processes were shown to form a complete system in which memes can reproduce.

Variation was explained as being caused by three mechanisms: cognitive mechanisms, transmission mechanisms, and production mechanisms. Cognitive mechanisms cause variations in memes due to differences in interpretation or due to random errors in processes such as encoding and recall. Section 3.4.2 contains a discussion of the implications of retention processes on meme variation, which seem to be a major source of variation but one that is not well understood. Transmission effects on variation are well-framed by examining them using the Shannon (1948) Information Theory, which provides a general framework for examining how errors are introduced into information during transmission. Variation due to production mechanisms is addressed in Section 3.4.4, where the influences of motor control are addressed. These three mechanisms seem sufficient to examine a significant amount of variation introduced into memes.

Selection pressure was explained as being caused by both cognitive and environmental factors. Information Theory provided a pivotal lens for examining environmental selection, since it provides the conceptual underpinnings for how the environment can limit the bandwidth of transmissions. Section 3.3 overviews the theoretical reasons that memes may undergo selection during transmission. The cognitive aspects of selection are the focus of the discussion of attention effects, examined in Section 3.4.1. As such, both cognitive and information-theoretic aspects are important for modeling selection pressure on memes.

As such, the synthesis of Bandura's (1986) observational learning processes and information theory appears to be complete with respect to representing meme evolution. This does not mean that all the subcomponents are well understood, however. This completeness extends to identifying the primary subsystems, but as with any system there are other systems beneath these that are not fully understood. This completeness indicates that the synthesis is sufficient to study memes.

Demonstration of Holism

The synthesis of Information Theory and Social Cognitive theory provides a holistic architecture for examining memes. As was explained in the demonstration of completeness, all three major elements of evolution require both theories in order to understand all the mechanisms involved in meme evolution. This is because memes require the concept of an agent, with a separation between the agent's cognition and the environment. Under this formulation, it is impossible to fully consider them in any architecture that cannot represent both an agent's internal cognition and the external environment for a meme. Since both theories are necessary for the model to be complete, this approach is a holistic architecture.

Demonstration of Workability

Implementing the conceptual model within a cognitive agent simulation was done to show workability. The intention of this implementation was to show that the conceptual model could be used as a basis for building a workable system for simulating memes and that this system could be applied to looking at useful situations. The PMFServ architecture was extended to include all the principal elements of observational learning (attention, retention, motivation, production) and all the elements of information theory transmission (source, transmitter, medium, receiver, destination).

The Stanford Prison Experiment simulation showed the ability of the implemented model to represent a real-world scenario that was grounded in empirical data and to explore an unresolved question about the ground truth. In this case, the question was the source of abuses such as throwing prisoners in "The Hole" and also the source of prisoner resistance. This simulation of the Stanford Prison Experiment showed that the PMFServ simulation could effectively simulate many of the key aspects of the real experiment, as shown in the external validity metrics presented in Section 7.4.1. Moreover, it showed the ability of the cognitive architecture implementation to compare different hypotheses about the origins of the ThrowInHole and Resist actions. While this exploration was not ultimately conclusive, the framework showed promise for looking at real world situations.

The Hamariyah Iraqi Village experiment demonstrated the utility of the model for examining meme competition. While based on a fictional village, the simulation examined a real problem that exists within peacekeeping efforts: the competition between peacekeepers soliciting information from the populace versus insurgents attempting to destabilize the region. These efforts were represented through the GiveInformation (Pro-US) and PlantIED (Pro-Insurgent) actions. The analysis of the Hamariyah simulation showed the ability of the architecture to examine meme competition and selection pressures. This analysis showed how personality factors that motivate agents to express certain memes and the social environment that predisposes them to learning them.

This working implementation includes all aspects of the synthetic architecture for memes. It also includes significant number of factors discovered during the literature review that have been found to mediate attention and motivation related to social information, such as the effects of authority, conformity, and novelty on observational learning. The implemented realization of this architecture was successful in representing and exploring both the Stanford Prison Experiment and the Hamariyah Iraqi village. As such, this demonstrates that the framework is workable and useful for examining memes.

Significance of the Systems Synthesis Model for Memes

This systems model of memes is significant, as to date no other model has applied this level of cognitive modeling to memes. The conceptual model created as the result of this research incorporates insight from dozens of studies and theories that provide insight into the mechanisms that affect meme evolution. The computational implementation incorporates extensive attention mechanisms and social influence mechanisms from the conceptual model, making it unique in its capabilities for modeling meme transmission and selection. Both the conceptual model and its computational implementation are pioneering work intended to stretch the bounds of meme modeling and analysis.

In the larger picture, the formulation of the conceptual model is the more important piece of work. The computational model, while useful, is only one possible implementation of the larger systems model for memes. This computational implementation was tailored to social learning of affordances, a specific meme. By computationally implementing different components of the conceptual model, simulations that focus on different cognitive and environmental effects could be created. In particular, the current computational implementation does not model mutation or other effects that create new variants of memes. By implementing additional components that focus on the transmission medium and learning effects, different types of analysis could be applied to memes. As such, much of the potential of the systems model for memes remains untapped. Additionally, through dialog with the larger scientific community it should be possible to extend and improve upon the systems model for memes. As such, this systems model for memes should help increase dialog about the factors that impact meme evolution.

8.2.3 Observability and Measurement of Memes

Chapter 4 explored how memes could be measured and studied empirically, based upon this new definition and systems model for examining memes. This exploration was intended to make a foothold for later analysis using the computational implementation. The significance of this work is that it demonstrates that memes, as defined in Chapter 2, are an observable and measurable phenomenon. This contribution explores two conditions that significantly affect the study of memes: how they change behavior (activate vs. inhibit) and if they are still reaching new carriers in the sample (reproducing vs. equilibrium). It also describes in detail how socially learned affordances can be studied as memes.

While this work presents a short examination of the methodologies used to select and examine memes in this research, this remains a very open area. This exploration did not address issues of measurement such as how to address memes with multiple vectors of expression (e.g. verbal and written, for example). This is an area where a meta-analysis of research from communications and semiotics might be valuable. Secondly, measuring memes in the face of variation is only briefly addressed in Section 4.1 and Appendix A. Since meme reproduction can introduce changes in the semantic information, different “subspecies” of a meme can emerge- posing problems for measurement. This is an issue that confronts researchers in the domains of linguistics and diffusion of ideas. As such, there could be value in aggregating insights from these fields and finding general approaches used for measuring memes in the face of variation. More research on measurement methodology for memes is an important research topic, which could extend empirical approaches to meme analysis.

8.2.4 Examination of the Stanford Prison Experiment Archival Data

As part of the Stanford Prison Experiment simulation design, a site visit was made to the Archives of the History of American Psychology to access the archived data for the Stanford Prison Experiment. Working on-site at the archives, a significant amount of data was de-identified and organized in electronic format from the original materials. These efforts to organize and aggregate the raw and intermediate data from the Stanford Prison Experiment may have historical value in terms of inventorying information from that experiment. This contribution may help later researchers who hope to work with the Stanford Prison Experiment holdings.

This thesis presents some of the information from the original data and papers in Section 6.1.1 and also in Appendix H. This information provides a quantitative and qualitative overview of the experiment. Reading these sections would be no substitute for reading the original articles (Haney et al., 1973a, 1973b) or any of Dr. Zimbardo's materials, such as "The Lucifer Effect" (Zimbardo, 2007). The authoritative articles are the best source for understanding the experiment. With that said, this research has focused on the subjects and baseline activities more than the original papers. As such, the discussion in Section 6.1 summarizes information that might only be gleaned by thoroughly researching the experiment or by accessing the holdings at AHAP.

8.3 Component Authoring

The next major contribution included authoring components for examining meme transmission and competition, as described in Chapter 5. The computational model built was consistent with the definition for memes defined in Chapter 2 and implemented a significant portion of the conceptual model for representing memes defined at the end of Chapter 3. The component authoring process explicitly specifies empirical and theoretical findings, forming the bridge between the available science and a maintainable computational implementation. This followed the component authoring process of implementing simple, atomic cognitive components. These components represented first principles that would affect meme transmission, such as selective attention and ingroup influence.

8.3.1 Cognitive Effects Operationalized

Table 8.1: Cognitive Effects Operationalized

Cognitive Relationship	Source	Specification Details
Max Attended Events	Cowan (2001)	∼4 Max Attended Items
Attention∼f(Novelty)	James (1890)	Positive Correlation
Attention∼f(Selection)	Simons and Chabris (1999)	Positive Correlation
Attention∼f(Motivation)	Fazio et al. (1994)	Positive Correlation
Recall∼f(Exposures)	Bornstein (1989)	Sigmoid-Type Equation ∼10 to 20 Exposures → 100% Recall
Recall∼f(Repetition)	Ebbinghaus (1913)	Sigmoid-Type Equation
Recall∼f(Exposures)	Ray and Sawyer (1971)	Positive Correlation
Persuasion∼f(Exposures) Persuasion∼ |f(Exposures)|	Ray and Sawyer (1971) Ray and Sawyer (1971)	Unclear Correlation Diminishing gains per exposure
Persuasion∼f(Social Influence)	Petty and Cacioppo (1986)	Positive Correlation
Social Influence∼f(Conformity)	Asch (1955)	Positive Correlation
Conformity=$e^{-4*e^{\frac{-S^{1.75}}{T}}}$	Tanford and Penrod (1984)	Explicit equation, s.t. S is # Sources, T is # Targets
Social Influence∼f(Similarity)	Platow et al. (2005)	Positive Correlation
Social Influence∼f(Valence)	Kelley (1955)	Positive Correlation
Social Influence∼f(Authority)	Milgram (2004)	Positive Correlation
Social Influence∼f(In-Group)	Tajfel (1982)	Positive Correlation
Social Influence∼f(Reference Group)	Kameda et al. (1997)	Positive Correlation
Social Influence∼f(Transferability)	Bandura (1986)	Positive Correlation

Section 5.2 describes how the cognitive effects noted in Chapter 3 were operationalized. While these components were implemented computationally for use with PMFServ, this contribution has greater value than simply enhancing a single cognitive architecture. Since a knowledge engineering approach was used to categorize and represent these components, this process operationalized knowledge from the original papers. In this context, operationalizing means that key insights provided by a scholarly paper have been reduced to mathematical relationships which can be represented formally. Operational descriptions form the bridge between underspecified social science findings and implementable computational models. An operational description of a finding or theory captures the necessary core relationships that must hold true for any formal implementation. Since this will frequently be underspecified, a large number of possible implementations can be based on such work. In this way, the operationalization of the these cognitive effects could be of value to cognitive modelers using other architectures. By examining the fundamental relationships

presented in Section 5.2, equivalent or alternative versions of these components could be implemented in other frameworks.

Table 8.1 summarizes the main operationalized cognitive relationships incorporated into this research, along with a summary of the relationship captured. For many of the correlation-type relationships, the impact of the factor with respect to attention or learning was estimated as explained in Section 5.2.5 and Appendix F where the attentional salience weights are explained. While these operationalizations remain underspecified, they provide meaningful insights into cognitive relationships that should be expected to affect attention, learning, and social influence. By steadily operationalizing findings about cognition, the expected behavior of a cognitive model can be specified with increasing accuracy. These expectations for cognitive models allow them to better simulate human cognition in a descriptive manner.

8.3.2 Attentional Salience Integrated Model

To integrate the operationalized findings, the attentional salience model was used to integrate these factors. As previously noted, the attentional salience calculation was built using the KISS (Keep it Simple Stupid) principle (Axelrod, 1997). Using a simple weighted linear sum, attentional salience was calculated using novelty, selective attention, motivated attention, conformity, similarity, valence, authority, ingroup influence, reference group influence, and transferability. This integration tied together disparate factors that would affect the transmission and selection of memes, while capturing the appropriate operationalized relationships. As such, this is a significant contribution in that it was the basis of a workable computational model for examining memes that was theoretically consistent.

However, using attentional salience to integrate disparate factors required significant assumptions that were made entirely for the sake of parsimony. Since attention is necessary for learning, factors known to impact learning were integrated using attentional salience as well. This prevented adding additional assumptions about how each factor affected attention and learning separately. While this approach satisfied the operationalized constraints, the form of equation does not make good intuitive sense. One would expect these factors to potentially impact attention, learning, and motivation in distinct ways. Instead, only attention makes full use of these cognitive factors and assumes that they are all linearly independent.

This simplistic integration of cognitive components must be considered the weakest part of the work accomplished, since little empirical data supported any particular combination of cognitive components. This weakness is a direct result of the weakness in empirical research that studies these factors, since studies do not typically consider more than one such factor at a time. In this respect, this integration identified significant gaps in the available science. The interaction of different cognitive effects have not been well integrated and their interactions are not well understood. Expanding empirical studies to consider multiple cognitive effects would be an important step toward understanding their larger role in society.

8.4 Meta-Model Library

The meta-model library phase consists of where maintainable and re-usable computational tools are implemented to help study social science problems. This research made significant contributions to this area. These contributions included implementing PMFServ cognitive component plug-ins, creating a Stanford Prison Experiment scenario for simulation, and modifying the Hamariyah scenario to examine meme competition. While these tools may be specific to PMFServ, they can be used to model, simulate, and study hypothetical and real-life situations. Additionally, the modified inversion count metric was implemented computationally and is a tool that may have broader applications than examining the order of meme transmission.

8.4.1 PMFServ Cognitive Component Plug-Ins

A major thrust in this research was extending the PMFServ cognitive architecture to enable it to represent meme reproduction and selection effects. This research built more than a dozen new cognitive components for the PMFServ architecture, as explored in Section 5.2. These components incorporate the operationalized cognitive effects, implemented to be consistent with the existing cognitive components in the PMFServ cognitive architecture.

The new PMFServ plug-ins modeled attention, attentional salience, associative learning, perceived novelty, selective attention (inattentional blindness), motivated attention, conformity, similarity, valence, authority, ingroup influence, reference group influence, and transferability. The internal validity tests in Section 7.3.1 demonstrated that these components work as intended and capture the key insights used to create them.

These components are re-usable and can be applied to other simulations and experiments. The majority of these components are also capable of working independently of each other, unless they directly use values from another component. This means it is possible to create an agent who only implements a learning model, for example. These components will advance research at the ACASA lab and other organizations using the PMFServ cognitive architecture.

8.4.2 Stanford Prison Scenario

The knowledge engineering for the Stanford Prison scenario was a considerable piece of the total work for this research. Each of the principle actions for this scenario were knowledge engineered, the typical schedule for the experiment was represented, and agents were initialized with personalities based upon their personality assessment data from the original experiment. This scenario now exists in the library of modeled situations within the ACASA Lab and can be analyzed to consider different thought experiments about the Stanford Prison Experiment.

8.4.3 Hamariyah Iraqi Village Modifications

This research also produced a streamlined variant of the original Hamariyah Iraqi village scenario. This variant reduced the total number of groups and agents, but introduced the capability for meme transmission within the village. While currently the Hamariyah scenario only includes the GiveInformation and PlantIED meme, it is straightforward to add additional memes to examine their diffusion and competition within the fictional village. This variant of the Hamariyah Iraqi village has been added to the meta-model library of PMFServ scenarios and can be applied to explore new research questions.

8.4.4 Modified Inversion Distance Algorithm

An additional tool implemented was the inversion distance algorithm used in this research. This algorithm improves upon typical inversion count algorithms in that it is normalized, so that it fits between 0 and 1, with 0.5 being the average distance between a list and its random permutation. The algorithm also adjusts for ties and right censored elements in either the test sequence or the ground truth sequence. Existing inversion distance algorithms researched were not able to handle censored elements, so this algorithm may have utility for other researchers examining sequences of events. To this author's knowledge, there is no openly documented algorithm for using inversion counts to calculate a normalized distance between sequences that can also accommodate ties, missing elements, and right censored elements.

This sequence distance metric was a useful extension of standard inversion distance approaches and may have value for other problems where censoring or ties cause standard approaches to fail. With further study of the distribution properties of this modified version of the inversion algorithm, it may also be possible to turn this distance metric into a null-hypothesis test. Since the distribution properties of inversion counts for random permutations were explored with some success in Margolius (2001), it may be possible to build off of this research to produce a null-hypothesis statistic for sequence comparison that is robust with respect to ties and censored data in both sequences.

8.5 Application Usage

The scenarios produced for this research produced meaningful research findings, which have helped to give insight into those situations and into more general scientific questions. The contributions from each experiment will be noted, as each simulation provided unique and useful findings of significance.

8.6 Stanford Prison Experiment Simulations

The Stanford Prison Experiment was examined primarily to determine if social learning played a significant role in the hostile environment, with a focus on guards throwing prisoners in "The Hole" and prisoner resistance against the guards. This analysis modeled these influences by modeling them as if agents learned something that made them recognize their ability to take the ThrowInHole and Resist actions (i.e. the affordances for those actions).

The simulations of the Stanford Prison Experiment showed that memes starting with certain innovators were a plausible mechanism for explaining the order that guards first initiated the action to throw a prisoner in "The Hole." However, they also showed the Full Knowledge condition was equally plausible, leaving the question unresolved. In the absence of clear signs that meme transmission significantly affected the order of first expressions, the Full Knowledge condition appeared to be the most likely case. These simulations and analysis of the data also showed that a demonstration of the meme by an authority figure was unlikely, given the available data. However, this should be taken with a grain of salt since the authority figure in the experiment did not have any personality modeling. Agents might have responded differently to an authority figure with a different personality, which would change how that simulation condition unfolded.

While not conclusive, this examination still presented a novel approach to examining memes within a complex environment. The diffusion analysis showed how situational and personality factors both influence how memes operate in a social environment. The Resist meme was shown to diffuse faster, since the prisoners were always on-site during the simulation and also because it had to be performed with greater frequency to accomplish the intended effect. The ThrowInHole meme was shown to diffuse slower and in phases, since learning was constrained by the limited interaction between different guard shifts.

Finally, the Stanford Prison simulation showed good correspondence with the external validity tests that were completed. These tests showed that on the majority of metrics chosen a priori, the simulation reproduced the expected results. This demonstrated that the approach for extracting data from the Stanford Prison experiment captured many of the important aspects of the scenario and might be useful for exploring other questions about the experiment.

8.7 Hamariyah Iraqi Village Simulations

The Hamariyah Iraqi scenario modeled the competition between the memes GiveInformation and PlantIED within an environment populated by agents who spend the majority of their day doing basic tasks in their village, such as going to work or going to the market to buy food. As such, these memes compete with each other and with the other actions available to agents in the scenario. These memes were implemented on a backdrop where members of the village would consider which one they would prefer to perform on a structure managed by US peacekeeping forces in an unstable region.

Examining each of these memes found that both of them could spread within the village, though expression of these memes was disproportionately associated with certain members in the scenario. The Hypothesis condition made reasonable assumptions about insurgents being aware of how to volunteer to plant an IED and police and government workers being aware of how to give information to the US forces. Under this condition, agents learned memes disproportionately according to their group memberships. Agents tended to learn from agents in the same ethnic group in this simulation. Considering Hamariyah was modeled to be a heavily factional atmosphere, this makes intuitive sense.

This simulation indicated that PlantIED spread more effectively and was expressed more commonly than GiveInformation. This was due to a few factors. Firstly, the larger groups in the scenario disliked the US. Secondly, most of the agents who avoided the PlantIED action also had personalities that led them to avoid the GiveInformation action as well, because both were dangerous. In short, the agents willing to risk their lives were more prone to violent actions rather than relationship-building actions.

This raises questions about the concept of a “hearts and minds” campaign. The problem may not be the hearts or the minds, it may simply be that the friendly and non-violent portions of the population would simply rather take the safest option available. This analysis indicated that providing security for the population would be a key factor for improving the expression of the GiveInformation meme. This finding is supported by some counter-insurgency analysts, who state that assuring security for the populace is the most important factor and necessary for a useful “hearts and minds” campaign (Krepinevich Jr, 2005). This simulation indicated that it might be possible to have agents who dislike the US overall still provide intelligence if they find it beneficial for their long term preferences, but only if their safety can be assured.

Finally, certain factors that seemed likely to influence meme expression were not found to be significant overall. Notably, employment level was not found to be a significant factor for volunteering to participate in IED activities. This is concordant with research such as Berman, Felter, and Shapiro (2009), who state that higher employment does not appear to decrease the likelihood of violent rebellion activities that result in civilian deaths.

This analysis showed that the model was effective for looking at competition of memes within the fictional Hamariyah Iraqi village. Using the model, it was possible to determine not only the effectiveness of each meme within the population but also the key identifying factors that determined which agents were adopting a particular meme. While this simulation was done on a fictional village, it is not hard to imagine using this model to examine competing memes within a real population. With appropriate data collection and knowledge engineering, this model could be a useful tool for studying competition between memes in a real life context.

8.8 Gaps In Science

Based on the body of research presented, this systems social science inquiry has provided helped bridge gaps in scientific knowledge but has also exposed new gaps that warrant scientific study. The contributions of this research on the gaps in social science knowledge will be discussed briefly.

8.8.1 Existing Gaps Addressed By Research

A few significant research questions were approached through this research. The major gap addressed by this research was to present one modeling approach that applies cognitive modeling to examining memes. This research has helped to integrate and organize the available science in order to build a systems model that can explain many of the factors driving meme reproduction, selection, and variation. Through this process, some corollary findings were discovered.

No Evidence that Memes Significantly Affected the Stanford Prison Experiment

As noted in the contributions from the Stanford Prison Experiment simulation, this research demonstrated that there was no specific evidence that memes played a major role in how the behavior of the guards and prisoners played out. While memes were not disconfirmed, they were not confirmed and therefore seem unlikely to be the sole mechanism that affected transmission. Even if ThrowInHole and Resist were passed as memes through the Stanford Prison

Experiment, it seems likely that the overall impact of such memes would be low. Since the subjects were in a confined situation (a prison), attention and guard shifts were the major factor limiting meme learning. After a dozen exposures, even an event of relatively low salience had a significant chance of being learned. Since Resist and ThrowInHole occurred with some frequency, everyone would eventually learn to perform the actions. This means that the main limiting factor was expression- who was willing to initiate those actions. As such, this analysis indicated that personality factors and interpersonal factors seemed to be the primary effects that affected behavior, rather than social learning effects.

Negative Feedback Loop Between Novelty and Attention

Another general relationship of interest is that in a primarily static, complex environment- novelty correlates negatively with attention. Stated in this way, this is a very counter-intuitive relationship. However, due to the effects of attention feedback, this finding probably generalizes to most situations. The reason for this negative correlation is that even if novel events tend to draw attention more, paying attention to them typically makes them less novel. This means that events that are only interesting due to their novelty will eventually become uninteresting, while events that are interesting due to other factors will retain their attention salience.

In the long run, the events and entities that remain novel are novel for a reason: they were uninteresting in other respects. For example, while a person might be more interested in watching new movies, if they have owned a new movie for five years without watching it then it probably means that there are other factors that make it less appealing. This correlation discovered in the data was unexpected, but appears to be robust and may be provable within the empirical realm.

Diffusion Rates Show a U-Curve

The diffusion rate analysis for both conditions showed a shallow U-shaped curve for the efficiency of diffusion of ideas. This is an interesting phenomenon. This implies that the early adopters and laggards tend to be easier to reach with memes, as compared to the majority. For the early adopters, this makes intuitive sense because the first expressions will capture the people who are most susceptible to the meme. However, it is interesting that following the initial dip, efficiency of exposures rises for the last agents. This is an interesting effect which could warrant further study. In general, the U-shaped efficiency curve for exposures is an interesting phenomenon and it would be interesting to test the conditions under which efficiency of exposures tends to follow this curve versus a linear or convex curve. These theoretical results could then be used to design an experiment to test if such dynamics are found in empirical data.

Situational Collinearity Limits Model Calibration and Validation

Finally, it was discovered during the internal validity analysis that even a simple, known model may be hard to infer from a limited data set. For example, the correlations shown between different social and attention characteristics were very different between the Stanford Prison simulation and the Hamariyah simulation. This meant that each simulation appeared to have a different attention model, if one only examined the Kendall correlations and logistic regression.

This is an interesting finding for the purposes of looking at social science factors in an experimental setting. Even if all factors work based upon fixed importance weights (as was the case for the attention model), collinearity between inputs could stymie trying to determine these weights. One implication for this finding is that standard regressions may be improper when estimating the influence of attention and social factors, due to the potential for multicollinearity. For this reason it might be more appropriate to analyze these factors with regressions that try to account for multicollinearity, such as the Shapley Value regression. The Shapley Value regression technique uses game theory to more equitably and stably estimate β weights (Lipovetsky & Conklin, 2001). The failure of an ordinary least squares regression to capture these factors in the simulation setting has provided useful insight into the fact that such an estimation technique would be inadvisable in an experimental setting for examining these variables.

This indicates that simulated experiments might provide value for determining the analytical techniques that would be robust in capturing important relationships. Cognitive modelers often attempt to look at complex problems and must operationalize relationships found through empirical research, which often contains gaps and ambiguities (Silverman et al., 2001). The attention model designed for this research was tested as if it was empirical data, through experiments. It might be possible to extend this approach to prototyping theoretical models, then using computational modeling to design experiments that could also be implemented in the real world to test these models. This would improve feedback between computational modeling and empirical research. Such improved connections would ensure that empirical research would also be directly implementable within a cognitive architecture, which could be used for theory-based hypothesis testing.

8.8.2 Gaps Exposed By Research

While this research has produced some interesting findings, it has also opened many new questions. From the empirical standpoint, there were many obstacles encountered with respect to operationalizing cognitive models. This is a typical struggle for cognitive modelers, since empirical research is not conducted with the intention to implement computational simulations (Silverman, 2004).

Units for Operationalized Constructs

A second major open question is the issue of operationalizing constructs that utilize ambiguous units. This is also a common problem with constructing cognitive models. When dealing with factors that have been measured according to ordinal relationships (e.g. high/low authority), it is quite likely that only a fraction of the total domain of the construct has been measured. This typically leads to situations where research presents how certain conditions alter the dependent variable by some magnitude, but give no knowledge of how great the magnitude of the input variable has changed.

While such relationships can be operationalized, the operationalized version can either be reduced to a simple correlation or else the operationalized version will be representing the relationship from the specific experiment. The second approach was used in this research when determining attentional salience weights. As such, the minimum and maximum values of certain computational constructs are actually representing those contained in the experiments. While it preserves the largest amount of information, such magnitudes ultimately are not capturing the true range of cognitive phenomena but are rather dealing with the limited range found in an experimental setting. Without being able to at least infer a global minimum or maximum level for such constructs, such an approach runs the risk of overstating or understating the true impact of certain factors.

This also has important repercussions with respect to how factors should be combined. This is a significant question, especially when considering how to integrate multiple related factors. In order to properly weight cognitive models, it is necessary to establish some form of generalizable units for constructs or to measure all the related constructs in the same context in order to understand how they interconnect. Clearly, the second approach is preferable but may be more difficult. On the converse, it might be possible to establish workable units for certain constructs that would give some insight into the relative importance of different contributing factors to the same phenomenon. Even if such units were coarse grained, they could greatly improve comparisons between factors. Such experimental results would significantly improve operational descriptions, without requiring grand-scale studies.

Methodology for Improving Upon KISS Models

Related to these obstacles, there are a number of open questions with regard to the implementation of the cognitive model used by meme-capable agents in PMFServ. The first question is if the factors that affect meme transmission can be combined in a more appropriate way. While this model for attention worked effectively for these simulations and captured some of the important relationships for learning memes, its assumptions were not entirely realistic. Part of the cause for collinearity between components to the attention model is that factors such as ingroups and valence probably do not work independently. While this implementation assumed that attentional salience is mediated by a linear combination of independent factors, realistically these factors should be modeled using a different and more complex system of governing equations.

This research applied a KISS approach to integrating different factors that affected meme transmission. With that said, integrating these factors according to a more advanced model (such as a variant of ELM (Petty & Cacioppo, 1986)) might improve the performance and quality of results. However, there are significant questions as to the appropriate methodology for selecting and validating a more advanced model. This is a major open question with respect to computational models of cognition. While the Keep it Simple (KISS) principle is simple to follow, the Keep it Descriptive (KIDS) principle is more complicated (Axelrod, 1997; Edmonds & Moss, 2005). We know the model should be more complicated, but neither the KIDS principle nor the available science gives solid guidelines to how to improve upon the KISS model.

Lack of Big-Picture Studies

Though the assumptions of the cognitive model oversimplify the situation, there is insufficient empirical research to defensibly select a better model. As such, the solution probably rests in better guidelines for when and how cognitive modelers can best collaborate with experimental researchers that specialize in empirical research. Ultimately, for cognitive modeling and social systems modeling to produce its best work, empirical research must clarify the ambiguities and gaps between sets of interacting factors. For this to occur, social systems modelers must explain the need and utility of big-picture studies.

Unfortunately, experimental researchers working on socio-cognitive phenomena do not typically design studies that consider large numbers of inter-related factors. For example, it seems surprising to this researcher that experimentalists have not spent more time measuring how all major factors of social influence interact. The groundbreaking research that identified authority, conformity, ingroups, reference groups, and the halo effect as significant to social influence is well over twenty years old. With that said, not one study could be found that measured half of these factors. Socio-cognitive research is an active field, with a variety of good and interesting research that detects new cognitive effects and relationships. If a subset of the field was informed of the needs of cognitive modelers in this area, there might be a greater interest in studies that examine larger numbers of factors simultaneously.

Cognitive modeling could be significantly improved by even a handful of such studies that give real insight into how such factors interact. Even if such studies could not be easily broken down into "statistically significant" chunks, a large number of analytical techniques exist for processing complex data. By having all the available data in hand, different potential cognitive models could be tested against such data using train-test approaches. While such studies might be more expensive than the typical study, the benefits to science would seem very meaningful.

Given the amount of expenditure used to analyze cognition using expensive equipment such as fMRI machines, it seems reasonable that an equally significant amount of money should be applied to figuring out how such cognitive constructs interact. What would the point be to understand where constructs activate in the brain, if one still had no idea how they interact as a group? While discovering and exploring cognitive processes is important, it is equally important to study how the known processes interact as parts of the same system.

Without such studies, cognitive modelers remain at a disadvantage for building and validating models. One major challenge of this work was the search to find an appropriate data set that could be used for validity testing. Ultimately, the Stanford Prison Experiment was used because it was the only experiment found that had detailed information about the personalities, context, behavior, and had the potential for meme transmission during the study. Similar issues are encountered when trying to improve and externally validate the cognitive components for the agent cognitive model- the data simply doesn't exist. With that said, having such data would be a major boost to cognitive modeling in general.

Effect of Memes on Culture Change

Finally, given the definition and conceptualization of memes presented in this research there is the open question of how memes fit into larger frameworks of society and culture. Giddens (1986) presents a view of society as being guided significantly by the interaction of agency and structure, such as rules. Memes appear to be a mechanism which can help model shifts within this structure. However, the ways in which social structures and rules affect memes have not been fully explored. The Tomasello (1999) ratchet effect view of cultural shifts is a second super-system model which has implications for the cumulative effect of meme evolution within culture. Super-system models of this nature provide important context for meme transmission, since they frame and constrain the macro-scale impact of memes within society. The larger implications of this sort of model within such a super-system of shifting incentives warrant further study.

8.9 Scientific Shifts

This research started out with the goal to produce a scientific shift in how memes are formulated and modeled. Based upon the theoretical and applied work completed, this goal appears to have been a success. This research is not the last word on memes, but instead will hopefully initiate new dialog about modeling memes.

8.9.1 Principal Scientific Progress

This work accomplished two major scientific shifts with respect to the study of memes. Firstly, this research connected memes with a large body of social science theory and empirical findings that are extremely important for understanding how memes evolve. Due to the origins and relative newness of memetics as a field of study, there has been a great unfilled need for such research (Heylighen, 1998). Even with this significant foray into examining memes in this research, more exploration is warranted to integrate cognitive science findings with memetics. With that said, this research provides new insights into how cognitive and environmental factors drive meme evolution. This research also provides new avenues for dialog about how these factors specifically affect memes within society.

Secondly, the systems model for memes was created and shown to be a useful model for studying meme evolution. Building off of a synthesis between the Social Cognitive Learning Theory (Bandura, 1986) and Shannon Information Theory (Shannon, 1948), this is a new approach to studying memes in that it attempts to richly capture the effects of cognition while still considering the effects of transmission and encoding. It is also novel in that it explicitly considers memes to be defined in relation to a population. The systems model was used to organize the empirical findings related to memes into a conceptual model for examining meme transmission. This model has shown itself capable of producing computational implementations that can be applied to real world questions, while keeping a rich conceptualization of memes. This model should help open new avenues of discussion and research for looking at memes within society.

8.9.2 Future Directions

There are a number of future directions for this research, following from the theoretical work and the computational work. Firstly, as noted previously in Section 8.2.1, the meme definition proposed in this work can be part of the larger discussion to develop a canonical definition for memes that is both ontologically adequate and supports empirical measurement. A second theoretical direction for this work is to delve deeper into the relationship between memetics and semiotics, in terms of scope and focus. In particular, it would be valuable to examine the relationship between memetic approaches to analysis as compared to semiotic approaches to similar topics.

Another major theoretical question is how meme mutation could be modeled in a generalizable and empirically-grounded manner. While the computational model implemented for this research was sufficient to model meme reproduction and selection, mutation mechanisms were not implemented. A significant reason for this design choice was that the mechanisms for mutation found during the literature review were typically specific to a particular medium, expression type, or memory encoding issue. While Information Theory covers most of these cases, the specific characteristics of the medium and noise must be modeled in order to examine a real world scenario. As such, it would be interesting to find literature that examines general, empirically validated effects that can help examine meme variation.

The future direction that could most greatly benefit this type of work would be richly detailed empirical experiments on cognitive effects, which take into account many inter-related constructs. While such experiments might be avoided because they are "messy" to interpret, cognition seems to be a rather messy business. To understand how different cognitive factors interact, such experiments are necessary. Moreover, the internal validity research noted in Section 7.3.4 and Appendix J.1 indicates that, by not measuring such factors, one might infer distorted relationships that are specific to the experimental context. Big-picture studies that look at many related variables would be a major step forward for understanding cognition, since these would give a glimpse into how many parts interact rather than a focused view of only two or three parts at a time. However, as observed in the difficulty of using a typical regression to estimate the contribution of such factors, it is important to consider the analytical tools being used to examine the parts.

Another avenue of further work would be to examine the larger implications of this model of memes on the social system. Much of this research has focused on how the social system affects meme reproduction and selection. The alternative question is also very important: how do memes affect the larger system as a whole? A modest application would be an implementation similar to the Hamariyah scenario, but working with real-world data rather than a fictional village. Such a study would be an excellent further external validity test and would also allow the model to provide insight into real issues of meme adoption and selection in a complex environment. The success of the model in such a context would be a significant step forward for analysis of viral marketing issues or even military strategy, such as the spread of a new IED design. In this way, this research can be applied to significant hands-on problems in the short term.

Finally, a future application for this research would be to model a real-world problem in tandem with an empirical experiment. By conducting an experiment and collecting appropriate data, it should be possible to better establish where the computational implementation of this model succeeds and fails for representing meme dynamics. A significant challenge to such research would be to collect the amount of data necessary to initialize PMFServ agents, which was one of the reasons why this design was not initially chosen to help validate the model. However, given that the computational model has passed its initial tests it may be valuable to apply it in tandem with empirical data collection which could help test and improve the cognitive model for memes.

In theory, such an approach could even be done as a complement to a big-picture study as mentioned earlier. As shown in the systems social science development cycle, the exchange of information between systems modelers and discipline specialists helps to advance science. By building big-picture experiments specifically designed to produce operationalizable results, in tandem with a cognitive modeling approach, a closer exchange of knowledge would be produced. This sort of approach may ultimately be the future of cognitive modeling– to tighten its connections with empirical research. For cognitive research on more limited phenomena, such as reaction time or certain visual attention tasks, this interconnection is already the state of practice. However, for models of social phenomena this tight coupling is uncommon. This social systems model for memes would be well-suited to pairing with an empirical equivalent, hopefully as part of a scientific shift in socio-cognitive modeling in general.

Appendices

Appendix A

Speciation of Memes

Similarity of memes lacks a definitive metric. The problem of identifying memes has correlates to the problem of identifying asexually reproducing species. Biologists classify organisms using three types of information: DNA, phenotype, and the ability to interbreed. Currently scientists classify organisms by genetic features, but historically they have been grouped by behavior and physical features. This change in paradigm has not caused sweeping changes in scientific classifications (Lewin, 1997).

By analogy, phenotypical similarity for memes provide a necessary indicator in defining the meme reproduction process. Similarity of memes can be measured behaviorally, by physical characteristics. Variants of a meme should exhibit physical similarities in behavior or resulting signs. The study of imitation in behavioral psychology follows this paradigm (Zentall, 2007).

Unfortunately (or perhaps fortunately), humans possess a variety of means to communicate the same information. The similarity of transmissions may be neither necessary nor sufficient, but provides a basic metric. A more advanced version of this analysis would involve developing sets of behaviors classified by their likelihood of expression after some sets of observation. For example, a meme promoting suicide bombing may be learned by watching a bomber, hearing about a bomber, or reading about a bomber. Agents perceiving this meme should be probabilistically more likely to engage in one or more of the transmission behaviors. A Hidden Markov Model (HMM) could be appropriate for this sort of analysis. Care must be taken applying such approaches however, because not all contagious behavior involves a meme. Appendix B goes into great detail on this matter.

Perception equivalence provides an alternative metric for similarity through semantic traits. Classes of indistinguishable stimuli provide the backbone for perceptual equivalence. The affordance theory of perception and attractor theory of neurological activation are consistent with this metric (E. J. Gibson & Pick, 2000; Mainzer, 2007). This metric implies that if memes are perceived similar, they are similar. This methodology allows a much broader range of behavior to be classified as the same meme, but depends greatly on whose perception is used for classification. This introduces confounds when dealing with a sufficiently heterogeneous population. Due to cultural grounding a researcher may classify behaviors along different lines than the target population, a problem of great attention in anthropology (Herskovitz, 1952). Such problems do not negate the usefulness of this approach, but do indicate a need for caution.

Neither metric for meme similarity holds the same convenience for analysis as DNA in genetics, but the existence of quantifiable similarity metrics allows for evolutionary analysis. Placing bounds on a similarity metric allow behaviors to be classified into a single meme. Classification grants the ability to measure meme proliferation and evolution.

While using similarity to define memes may seem overly relative, the same argument could be made for genes. Gene instances contain small mutations when compared to the parent. Even during human cell division, the telomere of chromosomes shorten after each replication- one of the proposed causes of aging (Tsuji, Ishiko, Takasaki, & Ikeda, 2002). Equivalency of gene sequences follows a similarity metric, requiring a certain level of differentiation before defining a new entity. Foregoing classification of memes would be akin to studying biology without a taxonomy- every organism wholly separate and ungeneralizable.

Appendix B

Imitation and Other Behavioral Transmission Mechanisms

Multiple mechanisms describe the content of learning, if any, that can explain the spread of behavior through a population. These mechanisms are listed in Table B.1. Humans possess the ability to learn using all these mechanisms, while other animals have been shown to pass knowledge using a subset (Zentall, 2007; Whiten et al., 1999). Complicating matters, behavior can spread through mechanisms like contagion that transmit no semantic information. Reflexive smiling and yawning fit this pattern. Other mechanisms pass semantic information, but indirectly related to a behavior. For example, stimulus enhancement is capable of transmitting a meme giving increased attentional salience to an object. Differences in objects and perception could result in evolution of which object receives increased attention through a population, without evolution of individual behaviors.

As can be noted, a number of mechanisms that allow behavior to spread are not not actually meme reproduction. Those marked as “Maybe” do indicate social learning, but the content of the learning does not consist of a meme to be expressed. Instead they pass basic environmental information which gives support to express a behavior. If this piece of knowledge is passed from one agent to another as a result of behavior affected by this information, it is a meme. The meme in this case is not an action, but a piece of environmental awareness. These would be memes that do not necessarily involve any imitation.

The uncertainty around discriminated following is analogous to the Chinese Room problem, which hinges on the meaning of understanding (Searle, 1980). This is because an agent can store and replicate a behavior, but with no knowledge of semantic information. Searle would say in this case that an agent does not know the meme in that case. However, the agent can still express the meme. The

Table B.1: Imitation Mechanisms (Mitchell, 1987)

Mechanism	Description	Meme?	Semantic Information
Mimicry	Genetically predisposed response where agent appears like another organism	No	None
Contagion	Genetically predisposed response to perform a certain action when observing signs of that action (ex. flocking)	No	None
Social Facilitation	Motivation changes due to presence of other agents	No	None, behavioral semantics rediscovered
Incentive Motivation	Motivation changes due to knowledge of results of a behavior	Maybe	Existence of a reinforcer (contextual information)
Local Enhancement	Attentional salience increases for areas where another agent acts	Maybe	Existence of a location (contextual information)
Stimulus Enhancement	Attentional salience increases for manipulated objects	Maybe	Existence of an object (contextual information)
Discriminated Following	Miming the product of a behavior (ex. matching sound pitch)	Unclear	None retained internally, but syntax can evolve.
Observational Conditioning	Pavlovian learning of a stimulus-outcome pairing	Maybe	Existence of a relationship between stimulus and outcome
Affordance Learning	Discovery of a new action in the environment	Yes	Action opportunity between agent and object
Behavior Imitation	Reproduce actions involved in a behavior, without feedback for matching	Yes	Relationship between action and outcome
Behavior Emulation	Reproduce outcome of an action using different capabilities	Yes	Relationship between action and outcome
Goal Imitation	Behavior to achieve a similar demonstrated goal	Yes	Relationship between motivation and outcome
Symbolic Imitation	Behavior demonstrating similar relationships as another agent's behavior but in a different medium or syntax (ex. language)	Yes	Relationships between syntactic elements

designation for discriminated following is ambiguous as a result.

True imitation is always a meme, but can take different forms. Individual differences between agents of different capabilities create systemic differences in behavior. For example, a taller person stoops to reach a shelf but a person half their height must stretch to reach the same item. Certain forms of social learning, particularly action learning, may operate under two distinct modes during this process. Imitation can occur between organisms of similar capabilities, through the use of mirror cells (Rizzolatti & Craighero, 2004). Systemic variations require emulation. Emulation achieves the same outcome in an environment, but with adapted behavior. Symbolic imitation is the most advanced form of expression, which maintains the relationships of behavior but through a different set of symbols. Transcribing spoken words is an example of symbolic imitation.

Appendix C

Effects on Expected Outcomes on Memes

C.1 Attribution of Control

The attribution of control for an agent determines their perceived level of control over their environment (Bandura, 1986). An internal attribution of control indicates that an agent feels very capable of changing their environment. An external attribution of control causes an agent to feel that external forces guide their life and they give very little input. Since attribution of control is a perceived level of control, it may not correspond with the realities of an agent's environment.

Psychological discussion of control frequently concentrates on control the most probable outcome, rather than the distribution. This conceptualization of control differs slightly from control over the distribution. The bulk of this discussion will define control as affecting the distribution of outcomes, with a brief note about control over expectation and most probable outcome.

Attribution of control can be generalized or specific. Generalized attributions about control correspond to psychological factors such as locus of control and self-efficacy. Locus of control ranges from internal to external, and indicate the amount that an agent believes their actions control the environment (Bandura, 1986). Individuals with learned helplessness often have external locuses of control, feeling that the environment holds almost full control (Alloy, Peterson, Abramson, & Seligman, 1984). Self-efficacy determines an agent's perceived ability to complete appropriate of actions to achieve goals (Bandura, 1986). Self-efficacy requires confidence in the ability to carry out certain behaviors and affect the environment, a superset of internal locus of control. For high internal attributions of control, an agent must believe they can control their behavior and that their behavior will control the environment. No research could be located linking generalized attributions of control to memes, but the potential exists that individual differences in these factors may effect meme spread at the societal level.

Attributions of control often to correspond to specific contexts and actions. Even individuals with significant feelings of helplessness may indicate confidence in certain settings. Attribution of control will be considered to be an agent's perceived ability to express a particular meme effectively.

Attribution mediates an agent's motivation by altering the level that an agent's influences the system. An external attribution of control may have a damping effect on the value of a behavior, reflected as apathy or inhibition (Barrowclough & Hooley, 2003). From a systems standpoint, this means an agent cannot drive the distribution of outcomes towards an intended outcome using that behavior. A meme expression which does not appear to invoke meaningful changes will lose magnitude for its motivator payoffs.

$$U(M) - U_0 = U_{Failure} + p((U_{Success} - U_{Failure}) + (U^* - U_0)) \tag{C.1}$$

The loss of value does not reflect a strict decrease, but a decrease in magnitude. Equation C.1 displays the difference in value between expressing a meme ($U(M)$) and not expressing it (U_0), with p representing the probability of driving the system to a given distribution of outcomes with utility U^* from a distribution of outcomes. $U_{Success}$ and $U_{Failure}$ represent any value independently assigned to successfully or unsuccessfully expressing the meme. Desire for self-efficacy should generally cause the success and failure terms to be positive and negative, respectively. However, no assurance exists that the value of expressing a meme will outweigh suppressing it.

p can be thought of as an agent's perception of control. At its maximum, the failure term drops out and only the success term and the outcome terms remain. Where the intrinsic value of successfully completing an action is small, this reduces to the change in value in outcomes. In this case, a meme with highly motivating outcomes will be attractive. However when the attribution of control is external, both sets of outcomes may be essentially irrelevant and only the potential failure term will matter.

If meme has poor outcomes, feeling powerless could make it more attractive. For example, a person with little sense of agency might feel little connection outcomes. Breakdowns in normal standards observed in the Milgram (2004) experiment and the bystander effect (Darley & Latane, 1968) are consistent with such a circumstance. In populations where agents lack much attribution of internal control, memes expression might be primarily guided by the path of least resistance and responsibility.

When attempting to control the most probable outcome, an agent may feel capable of generating change but still have poor self-efficacy. Under these circumstances, outcomes regain relevance because the baseline outcome differs from the failure outcome. Agents perceive their actions as having an effect, but not a useful effect. The implications of attribution of control under these circumstances becomes less clear.

Since memes must compete for expression, the specific distortions of each behavior's distribution outcomes could influence their fitness. Distortions may omit outcomes or distort their likelihoods. Strategies of minimal regret employ distortion by omission, ignoring all but the worst results of a decision. The distortion of probabilities could cause the perceived probability of outcomes to have a high magnitude of correlation with their expected results. If such a strategy is applied to a single meme, it will be reduced greatly in fitness. If such a distortion results from a generalized attribution, such as external locus of control, memes of least regret may be the most prevalent.

C.2 Probability and Uncertainty

The effects of probability and uncertainty on memes must be considered distinctly from the concept of control. Given a realistic perception of self-efficacy, some probability and uncertainty must exist when attempting to express a meme. Humans do not handle probability normatively or even consistently.

Prospect theory, the matching law, and risk preferences provide different perspectives on the handling of probability. Prospect theory provides an shape for a curve relating actual probabilities to perceived probabilities. The exact values of this curve have some covariance with the outcomes and circumstance. The prospect theory probability curve underestimates low probabilities, overestimates high probabilities, and rounds extremely low or high probabilities (Kahneman & Tversky, 2004).

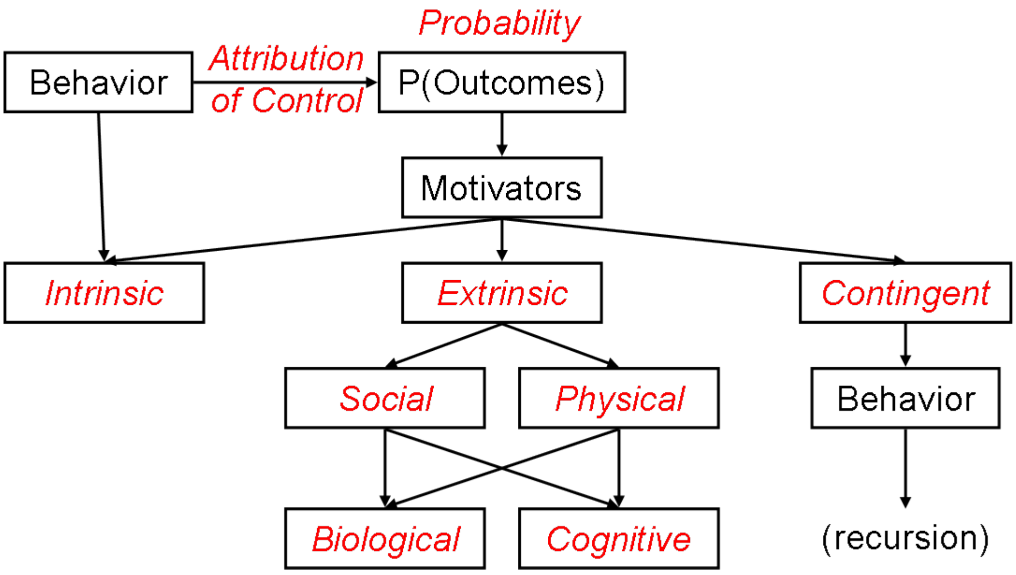
The matching law indicates a second handling of probability, or lack thereof. In the matching law, options have an appeal based upon the ratios of their expected benefits (Bradshaw, Szabadi, & Bevan, 1976). Under the matching law, probability collapses into the expected benefit and the behavior itself reflects the likelihoods. Reinforcement learning supports this form of outcome, where probabilities affect strength of acquired associations rather than explicit inferences (Sutton & Barto, 1998).

Explicit preferences about risky behavior could also play a role. If consideration of probability weighs outcomes, as is seen in utility theory, prospect theory would seem to apply. Where probability effects only play a role as situational cues, these might better be thought of as a special type of motivator. In this case an agent might consider intrinsic motivators for participating in risky behaviors, without weighing the extrinsic outcomes as a function of their likelihoods. Uncertainty tends to exert influence in this manner, with most people being averse to actions with uncertain outcomes (Epstein, 1999).

Appendix D

Effects on Payoffs of Motivators on Memes

Figure D.1: Decision Theory Mapping of Motivators (From Chapter 2)



D.1 Intrinsic and Extrinsic Motivators

Many motivational theories differentiate between intrinsic and extrinsic motivation (Malone & Lepper, 1987). Contradicting sets of intrinsic motivators have been proposed, as well as flat out denials of intrinsic motivation (Reiss, 2004). Second order conditioning experiments on motivator devaluation indicate that a portion of motivation becomes intrinsic to a trained behavior, implicating a role for intrinsic motivators (Colwill & Rescorla, 1990).

Intrinsic motivation means that behavior directly generates a psychological reward or penalty. Theorists such as Combs (1982) propose that intrinsic rewards build or maintain a self image. Extrinsic motivators provide benefit contingent on the environment's response to behavior. Reinforcers in instrumental learning are generally extrinsic, as these can be controlled. Intrinsic and extrinsic motivation are non-exclusive; most behaviors include elements of both.

It is important to look at motivators from the standpoint of the agent, not the experimenter. Intrinsic motivation exists when an agent is attracted to a behavior independently of its outcomes. Outcomes alter the value of the intrinsic component through feedback, but are not considered in terms of outcomes at the time of the decision. Intrinsic motivation could be considered as assigning value to being in a goal state, enjoying the chase if you will. For expressing memes, sure-bet extrinsic motivators will have very similar influences on behavior as intrinsic motivation. However, they could have very different valuation dynamics. For example an extrinsic motivator could be devalued independently of a behavior (ex. less value to money), while intrinsic motivation requires devaluing the behavior itself (ex. less value to work).

$$\mathbf{M_a}(Env) = \begin{pmatrix} \mathbf{M_a^{Intrinsic}} \\ \mathbf{M_a^{Extrinsic}}(Env) \end{pmatrix} \tag{D.1}$$

An expression for the total motivation M for an agent a in environment Env is posed in Eqn. D.1. Intrinsic motivation (M^I) means that some payoff exists for expressing the meme, regardless of external feedback. Memes with high levels of intrinsic motivation for reproduction can be expected to transmit in more contexts than those with primarily extrinsic motivation. This does not necessarily indicate that intrinsically appealing memes will be transmitted more frequently, but in more varied settings.

High extrinsic motivation should cause a meme to depend strongly on its environment and an agent's perceived ability to produce positive outcomes. Extrinsic motivation has been studied in detail by behavioral economists and psychologists alike, through with very different methods (Camerer, 2003). Extrinsic motivators reflect the perceived outcomes of how behavior will alter the state of the environment. Some of these outcomes will be perceived as rewards and punishments, while a vast majority of environmental changes will be simply ignored. Outcomes salient to an agent should drive behavior. Extrinsic motivators can be split into a multitude of subcategories, down to individual changes to the environment. The important subclasses of extrinsic motivators will be discussed in the following sections on physical-social and biological-cognitive motivators.

An important characteristic of extrinsic motivation is that the designation and valuation of an external motivator can be highly subjective. Cognitive bias research studies effects of this nature, which is given a good treatment in (Pohl, 2004). Reference point research has shown that an expected gain that falls through may be perceived as a loss. So then an agent may be motivated towards behavior to prevent losses of gains they have not yet received. Prospect theory indicates that acting to protect a loss of an equivalent amount will have greater utility than acting to produce a gain of an equivalent value (Kahneman & Tversky, 2004). While valuation of external motivators is subjective, ongoing work continues to expose these processes. Additionally, many situations allow for simplifying assumptions and still produce useful models. Economic models routinely apply methodologies such as rational choice and multi-attribute utility which explain considerable variance in valuing external motivators.

High intrinsic motivation might cause meme related behavior to be observed in contexts which seem either chaotic or universal. Intrinsic motivators rely on the environment to afford the action and nothing more. So while other memes might be changing value dynamically based on the environment, an intrinsically motivated meme keeps the same meaning at all times. This meme is expressed due to no other meme having a sufficiently positive set of external influences to exceed its value at that time. If the meme has very high intrinsic motivators, an agent might express it in every environment. Otherwise, the meme will crop up as a function of its alternatives. Since little direct correlation exists between highly intrinsic memes and the environment, any analysis of these will depend extensively on the competing behaviors presented.

Changes in meaning for each type should be reflective of learning processes. Intrinsic and extrinsic motivation reflect different types of information, the former assigning meaning directly to the behavior and the later assigning it through intermediary outcomes. Literature comparing stimulus-response (S-R) and response-outcome (R-O) learning addresses a similar distinction which may be connected to this question, as in Rescorla (1991). S-R learning assumes that an agent associates a certain environmental factor with an action. Agents remember the behavior on seeing the cue and feel drawn to express it without precisely knowing why. Habit-based behaviors fit this paradigm. R-O learning assumes that an agent has learned that a certain behavior produces certain outcomes. Agents will be drawn to the behavior as long as they are drawn to the outcomes. Intrinsic motivation and extrinsic motivation appear differentiated in a parallel fashion, one adding value to the behavior and the other to the outcomes. Methodologies employed by learning researchers in this domain could be employed to analyze memes as part of subject-based research. It also provides a body of literature which gives clues as to why certain memes might become intrinsically motivated or extrinsically motivated.

Altering extrinsic motivation for behavior follows a learning process which has been approached from many angles. Public policy, psychology, and marketing have explored the topic extensively. Incentive structures, operant conditioning, and persuasion respectively have input into these processes. Consensus opinions state that extrinsic motivation depends on a correlation, preferably causal, with a beneficial or negative event (Rescorla & Wagner, 1972). This association may be direct or through a network of associations. Certain events and stimuli already have value and the learning process defines how the behavior relates to an outcome.

For memes, extrinsic motivation indicates learned outcomes for expressing a meme within a context. Memes expression, like other behavior, will be expressed when environmental cues promote a perceived benefit to the behavior. Since extrinsic motivation depends on an outcome, devaluing the outcome should also devalue the meme. In this way, memes with high extrinsic motivation should respond to incentive structuring. Also, it should be possible to change the value of a meme by altering its connection to outcomes, such as by punishing an expression that was once rewarded.

Confounding the discussion, behaviors can be motivators for other behaviors (Premack, 1963). Two mechanisms explain this phenomenon. One explanation relies on contingencies. In this view a behavior does not motivate another behavior, instead the agent chains together sequences of behavior which are weighed based on their final outcomes (Townsend & Busemeyer, 1995). Decision tree approaches take this route, a common normative technique. The second view involves assigning a value or preferred frequency for a behavior directly, which can motivate other behaviors (Premack, 1963). Intrinsic value for a behavior falls under this category. Since the extrinsic and intrinsic motivations of contingent actions are considered, motivation has a form of limited recursion.

Developing an intrinsic motivator involves an association which changes the value of a behavior without any reliance on intermediate outcomes. Classical stimulus-response theories posit that environmental stimuli link directly to behavior (Dickinson, 1985). Assuming that no learning of outcomes occurs, pure S-R learning is intrinsic motivation for behavior. This means that intrinsic motivation may be acquired through conditioning (Cameron & Pierce, 1994). Intrinsic motivators may act as compensatory motivation, making a behavior desirable enough to reach its extrinsic payoffs. Attribution theory explores this concept, where extrinsic and intrinsic motivators work in a complementary fashion.

Terror tactics can be framed as a meme, following adoption characteristics (Weitz & Neal, 2007). Recognizing intrinsic motivators or the lack thereof could be very significant in determining the memes that will be adopted by insurgents at different levels. Zealotry recruiting relies upon a cycle of intrinsic motivation reinforced by completion of goals such as attacking enemies and defending against perceived victimization (Atran, 2003). However, the leaders of militant groups seek extrinsic payoffs such as economic aid or redress of symbolic grievances (Hafez, 2006). Understanding the salience of different motivators is a major step in understanding how memes will circulate.

D.2 Physical and Social Motivators

Extrinsic motivation can be split into physical and social components. Every behavior in society has two sets of consequences, physical consequences and social responses. Physical consequences act as a pure reaction to behavior and the environment. Physical events have beneficial or negative consequences, such as satiating hunger or causing pain. Many consequences rely upon the action of other agents, such as laws or courtship. Agents may also value social status or popularity, independently of physical outcomes. Bandura (1986) defines these non-physical reactions as "arbitrary" but in this context they will be termed as social. Since both physical and social events depend on consequences they may be represented as components of the extrinsic term for motivation, shown in Eqn. D.2.

$$\mathbf{M_a}(Env) = \begin{pmatrix} \mathbf{M_a^{Intrinsic}} \\ \mathbf{M_a^{Physical}}(Env) \\ \mathbf{M_a^{Social}}(Env) \end{pmatrix} \tag{D.2}$$

A physical outcome is any outcome relating to a change in the physical world, excluding psychological changes in other agents. Physical motivators include regulation of bodily processes and changes to the state of physical objects. Victory in total war may be considered a physical outcome. Physical motivators of memes depend on an agent's physical environment, valuation of outcomes, and self-assessed efficacy to produce outcomes. Physical outcomes may be probabilistic or poorly understood, but they are not goal-oriented. Memes with primarily physical motivation should vary based upon the utility of the concrete environment for their expression. A separation exists between the laws of physics and process of decision making for agents, dividing outcomes from the decision process.

A social outcome induces a change in the psychological state of another agent, such as a change of opinion or goals. Social motivators include physical reactions of other agents and regulation of social relationships. Victory in a war of attrition may be considered a social outcome. Identifying the social consequences of behavior often provides more insight than the direct physical results of behavior.

Social motivators create game-theoretic situations, due to other agents' action and beliefs. Game-type situations introduce equilibriums and behaviors based upon the perceived social context (Ullmann-Margalit, 1990). Expected social payoffs and punishments depend on the models an agent has for other agents, rather than physical systems.

Categorizing outcomes into physical and social bins has psychological and analytical appeal. Psychologically, humans perceive social outcomes differently than physical ones (Grassian, 1992). For example, being hungry due to a famine is perceived differently than due to being denied food. These differences lead to different emotional states, which have effects on motivation (Ortony et al., 1988). Analytically, this distinction helps identify which aspects of the environment are most essential for meme expression. Knowing when to influence the social environment versus the physical environment has significant policy implications.

D.3 Cognitive and Biological Motivators

Extrinsic motivation can also be categorized by the origin of value for a motivator. Being motivated by a biological need as opposed to a cognitive goal could provide significant differences in behavior. A meme expressed to satisfy hunger may be readily controlled, while one expressed due to religious views could be intractable. Consequences of action will only be motivating if they hook into goals and needs.

Biological motivators depend on sensation, either internal or external. Classical theories of motivation such as the (Hull, 1943) drive reduction theory have analyzed behavior using purely physical motivators. Since biological motivators depend on sensation, they depend on the current state rather than the anticipated state. For example, an animal does not eat with the anticipation of being full but instead eats to reduce its hunger. A biological motivator may be considered as an internal sensor that indicates a need to be filled.

From the standpoint of a meme, biological triggers provide a strong driving force to cue expression. For example, the Japanese expression "itadakimasu" is predictably expressed before eating. Depending on a meme causally to satisfy biological needs would make this connection even more direct. If agent must express a meme before eating, that meme will be very likely to spread when an agent is hungry. Memes conveying affordances for obtaining basic necessities should be expected to spread based on these patterns, such as chimpanzee communities learning to catch and eat ants using sticks (Whiten et al., 1999). Spread of courtship behavior and ideas could also have correlations to biological states.

A cognitive motivator consists of a goal or preference state for an agent. By assigning value to world states, an agent can work towards anticipated consequences. Cognitive motivators differ from biological needs as goals involve the creation and maintenance of semantic meaning, rather than being hardwired. For this reason, an agent can theoretically be motivated by almost any event or world state.

The consequences of a goal provide only part of its motivation. Regulation of self-efficacy generates a motivation to complete meaningless or even destructive goals. By completing a goal an agent affirms its ability to complete goals and exert agency over the environment, increasing perceived self-efficacy (Bandura, 1986). Goals then provide an intrinsic payoff when completed, for resolving the goal state. This is different than assigning intrinsic value to behavior, giving intrinsic value to a world state.

Commitment to a goal may implicitly increase the value of expected outcomes and its impact on self-efficacy perceptions. Sunk cost and endowment effects provide examples where increased commitment increases perceived value (Kahneman & Tversky, 2004). Cognitive dissonance can also create positive feedback between goals and motivation, with increased commitment forcing increased valuation.

Cognitive motivation should be expected to be less universal than biological motivation. While biological motivators will have almost universal appeal, the value of a particular goal or world state will vary culturally and individually. Cognitive motivators express and change the values of an agent. Identity is a key concept in examining values and beliefs , which can be leveraged to analyze expression of memes. From an expression standpoint, agents should be expected to express memes that draw them towards their internally preferred identity (Tajfel, 1982). They should also be expected to express memes which reinforce this identity among other members in society (Berger & Heath, 2007). The concept of identity will be examined more fully in the discussion of motivator valuation.

Appendix E

Implementation Paradigms

The exploration of memes in Chapter 2 approaches memes from a theory standpoint, attempting to synthesize the conceptual models from literature. Two approaches will be discussed to address the problems of implementing a model of this scope: model of models and separation of conceptual models from programmatic models. After noting the underlying paradigms of the implementation, the computational model will be explained.

E.1 Model of Models

Implementing a system for memes demands a model of models approach, since memes are an emergent phenomenon from underlying processes. Creating the model for examining memes requires two steps: laying out a composite implementation of cognition and setting up the environment. A well calibrated implementation of cognition should transfer to many different types of memes- not just memes that are not affordances. The environment must be set up for each scenario of interest, but generally involves physical models or other well determined systems. This means that implementing appropriate agent cognition allows moving the same agents to a variety of scenarios, with only a straightforward implementation of the environment necessary. This is one of the key advantages of agent based modeling. It is also an example of a model of models approach, where the simulated model involves a model of the environment and models of agent behavior.

Building a model implementation using sets of constituent models has significant advantages. Silverman (2010) describes these advantages in detail. The primary advantages of building a model out of submodels are flexibility, validity, and explainability. Emergence provides the process that macro level phenomenon may be reproduced using micro level models, the key to these advantages.

Without interacting subcomponents, emergence cannot occur by definition (Holland, 1998). For example, economic growth can be modeled using dynamic equations based purely on initial parameters. Alternatively, the underlying firms and economic institutions can be modeled and generate similar macro-economic parameters for analysis. Building a model of models can generate the same output dynamics as a macro level model, but with a better correspondence to the real phenomena being modeled (Lustick & Miodownik, 2009). Improved representation of the model through submodels generates advantages over implementing a single model.

A model of models has greater ability to build up from literature. Research tends to generate an assortment of disconnected models. Multiple models exist that explain the same phenomenon, with the similar inputs and outputs but different processes. Rather than tweaking ad-hoc parameters, entire models may be substituted that express a different proposed view of the same behavior. With the theory encapsulated in a submodel, a new version can be implemented for comparison of alternative models in a common framework. Using models of models offers the flexibility produced by object oriented programming techniques that foster reuse, subclassing, and common interfaces.

Employing submodels from literature provides two advantages. Firstly, each model can be based on a tested and peer reviewed research (Silverman et al., 2001). If literature produces new findings that update a submodel, an updated version of the submodel can be applied and tested independently rather than potentially invalidating the entire model. Submodels based on literature have the advantage of external validation done within the field of expertise. Secondly, the submodel can be tested to ensure that its characteristics match that of the source material. This means that each submodel can be independently examined for its correspondence with literature before attempting validation of the full model of models. If an implementation fails validation tests, this can expose assumptions made in the original model concept that may not have been stated. These effects combine to form positive feedback between computational models and conceptual models, strengthening both.

Finally, a tiered approach to modeling offers clearer explanations for events. An experimenter may drill down into the model to find parameters in the underlying theory that are pivotal to the macro model behavior (Silverman, 2010). These connections reveal new research opportunities for empirical study and theorizing (Lustick & Miodownik, 2009).

E.2 Separation of Concepts from Computation

Separation of conceptual models from the programmatic implementation offers distinct opportunities. Ideally models from literature could be easily implemented as distinct mathematical models and connected to form a system for simulating memes. In such an implementation, a conceptual model maps to an object in code. Realistically, operationalizing and connecting models from literature does not allow such convenience.

Confounds exist for implementing models from literature. A single model from literature may have ambiguous parameters definitions, unknown bounds, unstated assumptions, or unclear scales of measurement (Silverman et al., 2001). Ambiguities require assumptions on the modeler's part, attempting to reproduce the characteristics and intent of the model. While difficult, resolving ambiguities analyzes where the model may be unclear and ultimately informs literature (Silverman, 2004). Additionally, known conditions may exist where the model generates erroneous results. Depending on the purpose of the model, fixes for these problems could require an altered version of the conceptual model. Unless informed by newer literature or an expert in the field, attempting to correct a model's failures damages the model's utility in research. While the new model may be more accurate, disconnecting it from the original research makes it hard to compare expected and actual behavior (Silverman, 2004).

Resolving ambiguities and fixing problems related to a conceptual model in order to generate an implementation causes a new confound: the implemented model is better specified than the original. Multiple implementations could operate significantly different, even in meaning and number of inputs and outputs. Implementing a useful model can require testing multiple alternative realization following the same conceptual basis. For an independent model, this process allows comparison of implementations that ultimately inform literature. For example, a multitude of variants on prisoner's dilemma in game theory analyze the same payoff structure. Without such research, emergence of the tit-for-tat strategy would not have been discovered (Axelrod & Hamilton, 1981). For the total model, alternative versions genuinely contribute to research.

Within a model of models, alternative implementations raise the issue of interoperability. Conceptual models may share parameters or alter parameters of other models. For example, physiology influences attention based upon theories of motivated attention: a hungry agent will attend to food better. Which model should quantify hunger? Hunger could be calculated by the physiology, based on the stomach but the perceived hunger of an individual could vary greatly by their mood. A motivated attention model could calculate hunger, but this model would need to be changed to fit a new model of physiology. A separate hunger model could be created, but taking this approach in every case will explode the space of programmatic models to the size of the ontology. Deciding on the specifications of one model will guide the specifications of interconnected models.

For a single implementation with no expected adjustments, explicitly interweaving submodels gets the job done at the expense of flexibility. However, if every submodel hooks explicitly into the implementation of other submodels then attempting to add, remove, or rework a submodel could involve reworking the full model. Since the correct configuration of cognitive theories is not known, this approach is not well suited for modeling memes.

An alternative approach involves multiple types of computational models, which differ by their level of correspondence and emergence. Table E.1 notes types of implementations possible. A direct implementation realizes a conceptual model as a cohesive module in code. A composite implementation realizes a conceptual model from parts of multiple modules. While it may not be possible to directly correlate any cohesive piece of code to a composite implementation, all mechanisms and data for the model exist at each point in time. A metric implementation represents a conceptual model through its dynamics over time, such as a how a cellular automata can represent market equilibria. An emergent implementation does not exist at any given point in time, but is evident in the dynamics of the computational model over time.

Table E.1: Translating Conceptual Models to Code

	Time	
Correspondence	*Instantaneous*	*Dynamic*
Single Module	Direct	Metric
Many Modules	Composite	Emergent

In order to build a versatile model of memes, the theory discussed in Chapter 2 will be implemented by a combination of direct, composite, metric, and emergent methods. This allows the computational representation to lay out data structures and modules flexibly, while allowing the each theory to be identified and monitored within code.

This model of memes is designed to capture meme reproduction dynamics for learning of new actions, also known as affordances. These are an emergent phenomenon resulting from the interaction of agents and groups. Memes reproduce through an interaction of agent decisions and their environment. Meme expressions are represented using affordances of the environment, always present but not always known. This allows a meme expression to be implemented directly, based upon the action it represents. Other actions available in the environment are also important, as they provide competition for behavioral expression. These models of behavioral expression will be kept simple and are discussed for each specific simulation scenario.

As noted in Section E.1, the cognitive models are the centerpiece for the family of meme models. Agent decision making is implemented through a composite set of cognitive models. The cognitive models will be implemented through a mixture of methods, appropriate to the specific theories involved. Only a subset of theories from Chapter 2 will be implemented, tailored to capture specific dynamics. This will be the focus of the following section.

Appendix F

Attention Salience Weight Methodology

The attention salience weights for each factor determine the relative importance of that factor toward attention. These weights were determined by looking for empirical research that showed a relationship between the factor and the probability that an subject showed some learning from the stimulus, since that typically indicates that a significant level of attention was focused on the stimulus. The process used to select the papers was an informal version of the methodology used to mark up PMF models, as defined shown in Figure F.1 which is an section from Figure 4 in Silverman et al. (2001). This figure shows the fields that are considered when examining the suitability of a paper for helping to build a Performance Moderator Function.

The empirical articles examined focused on attention tasks, recall tasks, and influence tasks (i.e. the Asch conformity study). On the validity assessment scale, most of the articles rated as low or medium. That is to say, they either contained a conceptual model that was too vague to implement directly or which had some ambiguity about how to operationalize certain aspects. In particular, the biggest source of ambiguity across all papers was that the factors involved did not have well-defined units or ranges. These are complex conceptual issues, as operationalizing a concept such as authority is an ill defined problem.

For this reason, all factors used in the computational model assume that the min and max values of each factor are those found in the specific experiment used to model the factor. This allows a slope to be defined which estimates the amount of increase in the dependent variable as a function of the different conditions (ex. low authority, high authority). While this is clearly an oversimplification, it seemed more appropriate to model only the known range rather than make

Figure F.1: Performance Moderator Function Anthology Markup Fields. Reprinted from Silverman et al. (2001), p.7

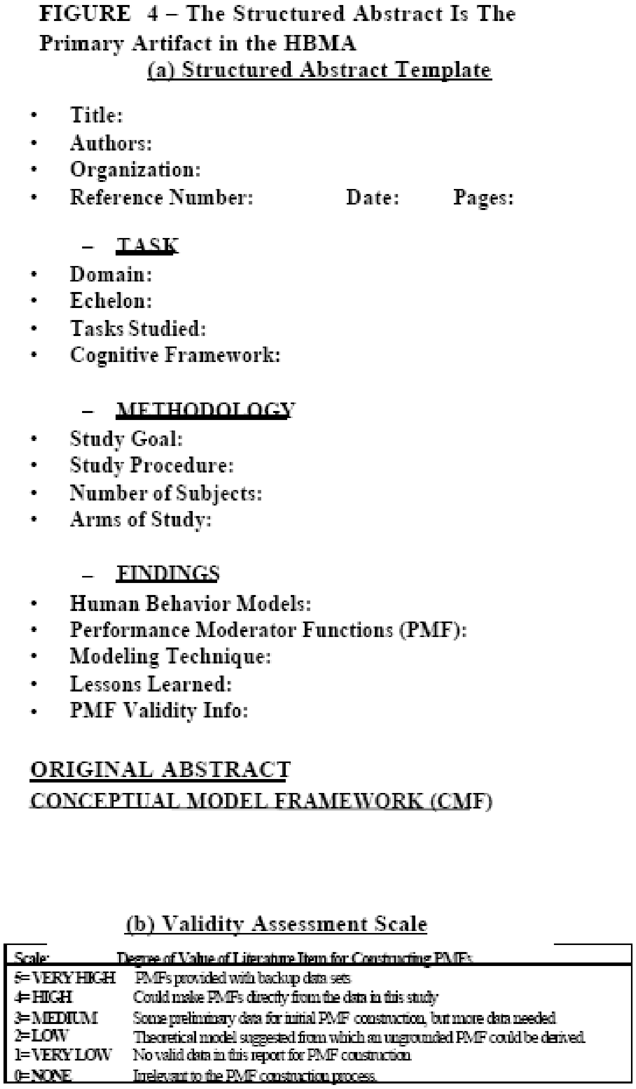


FIGURE 4 – The Structured Abstract Is The Primary Artifact in the HBMA

(a) Structured Abstract Template

- **Title:**
- **Authors:**
- **Organization:**
- **Reference Number:** **Date:** **Pages:**

– **TASK**

- **Domain:**
- **Echelon:**
- **Tasks Studied:**
- **Cognitive Framework:**

– **METHODOLOGY**

- **Study Goal:**
- **Study Procedure:**
- **Number of Subjects:**
- **Arms of Study:**

– **FINDINGS**

- **Human Behavior Models:**
- **Performance Moderator Functions (PMF):**
- **Modeling Technique:**
- **Lessons Learned:**
- **PMF Validity Info:**

ORIGINAL ABSTRACT

CONCEPTUAL MODEL FRAMEWORK (CMF)

(b) Validity Assessment Scale

Scale:	Degree of Value of Literature Item for Constructing PMFs
5= VERY HIGH	PMFs provided with backup data sets
4= HIGH	Could make PMFs directly from the data in this study
3= MEDIUM	Some preliminary data for initial PMF construction, but more data needed
2= LOW	Theoretical model suggested from which an ungrounded PMF could be derived.
1= VERY LOW	No valid data in this report for PMF construction.
0= NONE	Irrelevant to the PMF construction process.

assumptions about regions which have not been studied empirically. While it is likely that the authority of an post-doctoral experimenter is less than that of the President, the distance between those levels is unclear. This means that all weights for salience components are ultimately in terms of the experimental units. This means that relative importance of each weight for salience cannot be directly verified, because it is always possible that only a small range of any factor was measured and it might have a larger impact in other circumstances.

With these caveats stated, the process applied to the papers was intended to get the most out of the available information. The first step was to establish the input and output variables of interest. The second step was to determine the form of the empirical relationship, to the best of the experimenter's knowledge. The third step was to estimate amount that the input could affect the output, if known. Finally, each relationship was normalized so that the input variable ranged between 0 and 1. From these, the salience weights were defined. The goal of these salience weights was not to precisely duplicate the empirical attention or recall results but to get a best-guess estimate of the relative importance of each factor in orienting attention.

For example, Johnston et al. (1990) was used as the basis for estimating the importance of novelty. This article described a novel pop-out effect when a novel item was placed amongst familiar items, with subjects being more likely to remember the location of the novel word. For the PMF in this experiment, the input was the familiarity of a word and the output was the percent recall for the position of the word. The experimental relationship showed that a novel word had a recall percentage about 21% higher than a familiar word, in the novel pop-out condition (one novel word with three familiar words). For normalization purposes, it was assumed that the familiar words were maximally (novelty=0) familiar while the novel words were minimally familiar (novelty=1). So then, the salience weight assigned to the novelty factor was 0.21. This meant that a completely familiar word would have no increase in salience, while a completely new word would have a salience of at least 0.2.

This same process was performed for each of the noted factors to infer a salience weight for that input. This process gave an estimate of the relative importance of each of the factors, as compared to each other. This process was performed to help determine the relative importance of Number of exposures, Motivated Attention, Novelty, Selection, Authority, Conformity, Credibility, Ingroups, Reference groups, Similarity, Transferability, and Valence on the impact of a message to get an agent's attention and promote learning. Motivated attention, novelty, selection, authority, conformity, similarity, and valence were each estimated using these steps.

Ingroups, reference groups, and transferability were defined only as theoretical relationships and a literature search did not turn up a solid magnitude of impact of these factors. For example, literature shows that ingroups have higher social influence (Tajfel, 1982) and literature shows that social influences affect attention (Bandura, 1986) but empirical data on the strength of this impact is unclear. While they are stated as increasing social influence and imitation, they failed to pass the third step where a magnitude could be established to determine the impact of these factors. As a result, these factors were weighted based upon qualitative statements about their importance. This meant that transferability was given a low weight, as it was not indicated as a primary influence on attention (Bandura, 1986). Ingroup and reference group influence have been shown to have a significant impact on persuasion, but an operationalizable form of this relationship was not found. As such, these factors were given medium weight magnitudes comparable to other social factors such as valence and authority.

The number of exposures and credibility were processed using this methodology, but were found to not have a reliable relationship with attention or persuasion on a per-event basis. The number of exposures, while cummulatively influential, shows no reliable benefit for attending to later exposures of the same thing (Ray et al., 1971). The effect of credibility on social influence appears to be moderated by other factors and does not have a consistent impact on attention, if it has any impact at all. For this reason, these two factors were eliminated as salience factors and not included in calculating attentional salience.

Appendix G

Scenario Data Specification

The ideal data did exist for implementing even this limited model for affordance learning. The full data required to build a scenario within this model requires weighted social linkages, detailed personality assessment, pertinent situational factors, cultural information, and a measurement both of meme awareness and expression. For analysis of dynamics this information has to be collected over time, a significant undertaking.

Though full data does not exist, it was important to start with the ideal data. This allowed determining the strengths and weaknesses of the data available for modeling and experimentation. Weakly known parameters must be varied more than ones known with full confidence, for example. Table G.1 displays the full information for analysis running this model, broken into categories. Note that more data could be readily employed by this model by adding new motivators or social sub-models. However this level of data already exceeds what is available empirically, which indicates that increased empirical research could greatly benefit modeling of memes.

These data guidelines were used as a guideline in choosing scenarios Each PMFServ scenario directly uses many of these parameters.

Table G.1: Agent Data Specification for Studying Affordance Transmission

Situational	Alternate actions		
	Meme	Awareness	
		Adoption	
		Expression mechanisms	
	Material means	Employment	
		Money	
		Tools	
	Physical state	Disabilities	
		Physiology	Fatigue
			Hunger
			Stress
		Health	
		Skills	
Personality	GLOBE (House 2004)	Scope of Doing	
		Sensitivity to Life	
		Time Horizon	
	Hermann (2003) Standards	Ingroup Bias	
		Need for Control	
		Need for Openness	
		Need for Power	
		Task/Relationship	
	Valuation Prefs	Material things	
		Sacred things	
		Symbolic things	
Society	Relationships	Authority/Power	
		Credibility/Trust	
		Interaction	Context
			Duration
			Frequency
		Similarity	
		Valence/Liking	
	Groups	Structure	Authority
			Members
			Roles
		Interactions	Outgroups
			Reference Groups

Appendix H

Stanford Prison Experiment Data

The GSP Tree in the Stanford Prison scenario (and also Hamariyah) consists of numerous personality factors, as mentioned in the main text. Table H.1 below notes the meanings of each of the personality nodes. The dash indentations represent the hierarchal structure of the nodes within the tree. The origins of many of these nodes are noted in Table G.1 in Appendix G. This is available as a reference for interpreting references to these nodes within Section 6.1.3 on the Stanford Prison design and also for the classifications performed on the Iraqi Experiment in Section 7.5.3.

H.1 Stanford Prison Experiment Scenario Design Methodology

Representing the Stanford Prison Experiment is complicated by the issue of entry and exit of participants. Guard presence is complicated due to the fact that certain guards tended to stay later after their shift than others, based upon their level of comfort in being guards. Nice guards and avoidant guards often to left their shift more promptly than mean guards. Additionally, some exceptions were made with respect to guard shifts- including an overlap period caused by the experimenters requesting a double shift on duty. The first problem with guard shift endpoints was handled by giving guards actions for starting and stopping their shift. This allowed the guard agents to decide individually how long they wanted to stay after their shift and how early (or late) they would show up. The second problem of experimenter manipulation was more complicated since it involved inputs not simulated: the researchers. Since shift manipulation occurred as a response to an attempted prison break, this event cannot be treated as entirely exogenous. However, since it would be infeasible to model the researchers,

Table H.1: GSP Personality Factors

NodeName	**Description**
Goals Nodes	Short term goals, which connect to joy and distress
Individual	Overall Importance of Individual goals, e.g. Maslow (1943) hierarchy
- Belonging	Importance of feeling socially accepted and situated
- Esteem	Importance of feeling self-efficate and respected
- Physiology	Importance of basic needs, such as eating and sleeping
- Safety	Importance of personal safety and well-being
Standards Nodes	Standards of behavior, which define how an agent prefers to accomplish things. Connects to Pride and Shame.
Conformity_Assertiveness	Overall importance of conformity and individuality matters
- Assert_Individuality	Importance of expressing individuality
- Conform_to_Society	Importance of conforming to culture
- Respect_Authority	Importance of respecting authority figures
Exercise_of_Power_n_Culture	Overall importance of power balances in actions
- Be_Controlling	Importance of controlling others, using power
- Be_Open	Importance of being open to others, allowing freedom
Honesty	Overall importance of honesty and promises
- Keep_Ones_Word	Importance of keeping promises, being honest
- Use_Duplicity	Importance of lying for its own sake
Humanitarian_Sensitivity_to_n_Respect4_Life	Overall importance of considering lives and showing respect for life
- Life_Res_r_Sensitive	Importance of respecting and being sensitive to lives of others
- None_r_Sensitive	Importance of disregarding and being insensitive to others' lives
Military_Doctrine	Importance of adhering to military codes
- Shun_Violence	Importance of avoiding violence
- Use_Asymmetric_Attacks	Importance of attacking unevenly, even unfairly
- Use_Conventional_Attacks	Importance of using force-on-force conventional tactics
Scope_of_Doing_Good	Overall importance of doing good to others
- Bring_About_Greater_Good	Importance of good in the world, in general
- Look_After_Narrower_Interests	Importance of only looking after one's own
Task_Relationship_Balance	Importance of balancing tasks and relationships
- Be_Task_Focused	Importance of concentrating on tasks only
- Be_Relationship_Focused	Importance of building relationships
Treatment_Of_Outgroups	Importance of interacting with outgroups
- Outgroups_Are_Legitimate_Targets	Importance of targeting outgroups for discrimination
– Enemy_Is_Outgroup	Importance of targeting one's enemies
– Friend_is_Out_Group	Importance of targeting one's friends
– Neutral_is_Out_Group	Importance of targeting neutral parties
- Treat_with_Fairness_n_Justice	Importance of treating everyone equally
Preferences Nodes	Long term wishes for the world state. Connect to Like and Dislike emotions.
Desirable Future	Importance of good outcomes, by scope
-For_Everybody	Importance of everyone doing well
-For_the_Group	Importance of one's immediate group(s) doing well
-For_the_Self	Importance of self doing well
People	Importance of people doing well, by relationship
-Enemy_Faction	Importance of enemy factions doing well
-Friendly_Faction	Importance of friendly factions doing well
-Own_People	Importance of one's own members doing well
-Other_Groups	Importance of other neutral groups doing well
Places_n_Things	Importance of objects in the world
- Materialistic	Physical and monetary objects' importance
- Symbolistic	Importance of symbols, principles being maintained
- Wholistic_Spiritualistic	Importance of religious or spiritual matters

there is no easy solution to this issue. The modeling choice was made to assume that shifts ran regularly, rather than enforcing exceptions that might never have happened under different counter-factual scenarios.

The problem of prisoners being dismissed has similar implications. Prisoners were released early from the experiment as a result of their stress and emotional state- dynamic properties that are simulated. Based upon these factors, the experimenters chose when to dismiss participants. As with shift manipulation, it would be foolish to assume that prisoners would always have been released at certain times- even if they were in a very different emotional state due to different initial conditions. In this case, the Experimenter agent was allowed to take actions to release subjects from the experiment if their stress level was very high and they demanded release. This was implemented in the form of a conservative utility-based decision rule, since the experimenters were not modeled.

Appendix I

Modified Inversion Algorithm: Additional Information

Inversion counts are a metric for measuring the distance between orderings of different sequences. This class of distances is commonly used in genome research. The algorithm designed for this research was built to be a one-way test, comparing one sequence for its fit against a ground-truth sequence. It compares these sequences based upon the number of single-element moves required to turn the test sequence into the ground truth sequence.

This algorithm differs from a standard inversion count in a few meaningful ways. Firstly, this metric has been adapted to account for ties and right censored elements, both in the test sequence and in the ground truth sequence. Secondly, this metric is normalized- it is guaranteed to fall between 0 and 1 so long as it can produce a value. Finally, this metric is typically unbiased in that the distance of a random permutation of a sequence will have a distance of 0.5. The only bias metric is evident if averaging over tests where the algorithm cannot establish any distance, due to too many right censored ground truth elements. A correction could be can to these conditions, but if the metric is consistently not producing a value then it should not be used. The algorithm was implemented in Python and the source code may be made available at a future date.

The process by which the algorithm determines how close a test sequence is to a ground truth sequence goes as follows:

1. Replace elements of test sequence into their sort-order numbers, based upon their position in the ground truth sequence.

2. For all ties in the test sequence, count the maximum number of inversions that could occur within the tied subsequence. Sum these inversion counts and set aside.

3. Sort all tied subsequences so they will have no inversions.

4. Flatten the test sequence, evaluating all ties into elements positioned in their sorted order

5. Count the number of inversions for the flattened sequence, using a merge sort algorithm

6. Calculate maximal number of inversions for the flattened sequence

7. Subtract the inversion counts from ties from the ties from the maximal number of inversions, i.e. maxInversions -= sum(tie inversions)

8. If the maximal number of inversions is greater than 0, return the number of inversions divided by the maximum inversions, i.e. Inversions/MaxInversions

Right censored elements for the test list are treated as if they are one big set of ties at the end of the sequence, since ties and right censored elements are handled in the same way (count how many inversions they could cause, sort them correctly, and subtract those possible inversions from the maximum possible).

This algorithm is unbiased so long as the ground truth sequence is well formed. A well formed ground truth sequence is one where all permutations are assured of producing a result with at least one possible inversion. The algorithm shows a bias if one blindly excludes the cases where the maximal number of inversions is zero. This is because this increases the chance that any list with at least one possible inversion will perform better than chance.

For the ground truth sequences in this research, the algorithm is unbiased. In the future, work may be performed to compensate for biasing due to excluded cases by detecting the possibility of these cases and allowing an option for an adjusted distance score that will be unbiased for all sequences.

Appendix J

Supplementary Data Analysis

This section contains additional analysis which is not included in the main document for the sake of brevity and focus. These additional analyses were run and may be of interest. A few notes on each will be presented, along with the figures and tables.

J.1 Event Salience Component Weights (Simulation)

To look deeper in to the issues of collinearity noted in Section 7.3.2, correlation matrices were calculated from the observations data set. Table J.2 shows the Kendall correlation calculated for the Iraqi village scenario under the random start condition. This correlation analysis looked at agent attention and learning from all events and not just those involving a meme, since they are processed by the same cognitive models. For space considerations, the variables were assigned labels designating them as the dependent factors (Y_i) and the input factors (X_i). Agents were allowed to learn from everything they attended, so attention and learning have a correlation of 1.0 (omitted from the table due to space constraints). Table J.1 shows the equivalent matrix for the Stanford case. The Pearson correlation matrices are similar to these and are examined in J.2.

When looking at these matrices, it must be kept in mind that each has over two million samples- so all correlations are technically statistically significant (p<0.0001). Moreover, these factors were set up to be primarily orthogonal so the emergence of significant correlations between them is an interesting phenomenon. For the purposes of this analysis, anything with a correlation magnitude less than 0.05 will be considered uncorrelated. While there may be some small interaction, these seem unlikely to be of major importance. Those between 0.05 and 0.1 will be considered marginally significant. These will not generally be discussed unless

Table J.1: Correlation Matrix for Stanford Prison Attention (Hypothesis Cond.)

	Y	X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Attended (Y)	1.00	-	-	-	-	-	-	-	-	-	-
Novelty (X_0)	-0.06	1.00	-	-	-	-	-	-	-	-	-
Motivation (X_1)	0.00	-0.03	1.00	-	-	-	-	-	-	-	-
Selection (X_2)	**0.17**	-0.02	-0.07	1.00	-	-	-	-	-	-	-
Authority (X_3)	0.01	**-0.12**	-0.05	-0.06	1.00	-	-	-	-	-	-
Conformity (X_4)	-0.06	0.01	0.09	**-0.20**	-0.05	1.00	-	-	-	-	-
Similarity (X_5)	**0.12**	-0.07	0.02	**0.16**	-0.01	**-0.12**	1.00	-	-	-	-
Transferability (X_6)	**0.13**	**-0.27**	0.08	-0.02	0.07	-0.07	0.08	1.00	-	-	-
Valence (X_7)	**0.15**	**-0.17**	0.07	**0.20**	-0.01	-0.03	**0.35**	**0.15**	1.00	-	-
InGroup (X_8)	**0.15**	**-0.26**	**0.22**	0.01	**-0.10**	0.07	0.06	**0.70**	**0.23**	1.00	-
Ref Group (X_9)	0.03	0.04	0.07	0.06	-0.04	**-0.10**	**0.29**	-0.03	0.04	-0.02	1.00

Table J.2: Correlation Matrix for Hamariyah Attention (Randomized Cond.)

	Y	X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Attended (Y)	1.00	-	-	-	-	-	-	-	-	-	-
Novelty (X_0)	-0.05	1.00	-	-	-	-	-	-	-	-	-
Motivation (X_1)	**-0.17**	**0.10**	1.00	-	-	-	-	-	-	-	-
Selection (X_2)	0.03	-0.03	0.02	1.00	-	-	-	-	-	-	-
Authority (X_3)	0.01	-0.02	0.02	-0.00	1.00	-	-	-	-	-	-
Conformity (X_4)	**0.10**	-0.01	**-0.20**	0.01	-0.07	1.00	-	-	-	-	-
Similarity (X_5)	0.00	0.02	**0.20**	**0.19**	-0.01	-0.03	1.00	-	-	-	-
Transferability (X_6)	0.01	-0.05	**-0.14**	**0.10**	-0.04	**0.39**	0.00	1.00	-	-	-
Valence (X_7)	**0.10**	**-0.17**	-0.07	**0.19**	**0.08**	-0.02	-0.01	0.01	1.00	-	-
InGroup (X_8)	0.07	**-0.10**	0.03	**0.14**	-0.01	-0.02	0.07	0.04	**0.28**	1.00	-
Ref Group (X_9)	0.07	-0.06	-0.00	0.09	0.01	-0.04	0.06	0.00	**0.23**	**0.68**	1.00

they are part of a larger pattern, but they do indicate some level of interaction between the variables. Finally, anything with a correlation stronger than 0.1 will be considered significant and anything higher than 0.2 is highly significant.

J.1.1 Generalized Correlation Trends (Both Simulations)

Using these standards, the correlations shared by both simulations will be discussed. The first factor, Novelty, shows a very interesting effect. In short, novelty correlates negatively with almost everything else. In particular, it correlates negatively with ingroups and valence in both conditions. This results from a de-facto negative feedback loop. While novelty increases salience, attending to certain events increases familiarity with them. This means that while novelty may initially correlate positively with attention, over time it will tend to correlate with events that are otherwise uninteresting. Part of the reason for this interaction is that each simulation uses a fixed set of actions and actors-very little new stimuli appear. As a result, agents and actions that remain more novel are the ones that had low salience due to other reasons. This indicates that while novel stimuli may be more salient than other stimuli, unfamiliar stimuli in a fixed environment will tend to have low salience in other respects.

Selective Attention is expected to vary based upon the simulation, since it depends upon agent's actions within their environment in who they choose to interact with. With that said, in both simulations it correlates positively with valence and similarity. This indicates that agents tend to perceive and interact more with agents that they like and who have similar personalities. While this may not hold true in all contexts, it correlates with findings from network science that indicate that increased similarity correlates with increased interactions (Christakis & Fowler, 2008). However, it is not difficult to imagine contexts where dissimilar rivals would be the focus of selective attention- such as in a war or other competitive environment.

Finally, Ingroups and Valence are correlated in both cases. This makes intuitive sense, in that it would be expected that people sharing an ingroup would like each other more than people in outgroups. This dynamic matches with a significant body of research on social identity theories, which posit a preference towards people who appear to be in the same ingroup (Tajfel, 1982).

While each of these correlations match with intuition, none of them were directly coded into the cognitive model of the agents. Each of these correlations is emergent from the perceptions and decisions driven by their cognitive models. Of these, the novelty effect is quite interesting. This model predicts that in the long haul, even if novelty increases attention salience, novel stimuli will eventually tend to correlate with less salient aspects of an agents environment if they are free to explore it.

J.1.2 Simulation-Specific Correlation Trends

The remaining correlation effects were only significant in one simulation, but in some cases were highly significant in that simulation. These are interactions that are due to the specific construction of the simulation. Additional analysis of each model was performed to examine why each simulation evidenced its particular correlations. Novelty correlates negatively with transferability, but it correlates much more strongly in the Hamariyah Iraqi village. This is due to the design of the simulations. Since the Hamariyah Iraqi village is much larger, many actions are only possible in certain locations. Conversely, the Stanford Prison experiment's actions are typically available to all agents when they are perceived. Since agents are less likely to go places where they can't perform any actions, actions that a Hamariyah agent can't do will tend to be more novel.

Motivation was the most complicated of the factors, which had significant correlations in both simulations but was not consistent between them. A correlation with motivation indicates that agents would most like to have the results of events that certain agents are taking. Motivation is complicated because it depends highly on both the environment and the GSP personality model. Since motivation is moderated by the GSP tree and the GSP tree helps define actions, one would expect that similarity would correlate positively with motivation. In Hamariyah, a fairly strong correlation exists between these factors. However, this correlation is barely present in the Stanford simulation. Instead, motivated attention is correlated highly with membership in the same ingroup. Despite having a high salience component weight, motivated attention was the worst indicator for attention. It correlated moderately or highly with a large number of factors in both simulations. This may indicate that motivated attention should not typically be modeled separately from such factors, but should be a function of such factors.

Selective Attention shows a significant negative correlation with conformity influence in the Stanford simulation. This is due to attentional limitations. Since all events occur simultaneously and a maximum of 4 may be attended at any one time, periods of high conformity can result in periods where all attended events match conforming action but there are still a large number of unattended events with the conforming action. As seen in the Hamariyah condition, this effect disappears when each event is examined individually.

Authority had a small negative correlation with ingroup membership in the Stanford scenario. This is readily explainable due to the power differential between groups. Prisoners observed a larger total number of the events in the Stanford scenario than guards, since guards were on shift only part of the time. Since prisoners typically had lower authority than guards, this biases the observations so that authority correlates negatively with being in the same group. The Hamariyah scenario had no such power differential, so this effect disappears in that scenario.

In the Stanford scenario, Conformity correlates negatively with Similarity and membership in a Reference Group. This seems to indicate that the more central members of the simulation are less prone to acting at the same time as other agents. These effects do not carry over to the Hamariyah scenario and appear to be incidental. In Hamariyah, Conformity correlates with Transferability. This results from the fact that the most commonly occurring actions in the village are also the ones that are most commonly available. As with novelty, it is a structural factor due to the situation. In the Stanford scenario, the majority of actions are physically possible at all times so the frequency of actions is mainly guided by personality and interpersonal dynamics. Conversely, the village scenario's larger scope causes the ability to perform actions to be much more context dependent. This means that the transferability of actions is a proxy for their general availability, which affects the ability of large numbers of agents to simultaneously engage in an action.

The Stanford scenario shows a strong positive correlation between Similarity and Valence, as well as between Similarity and membership in a group with a higher reference value. These factors correlate much more weakly in the Hamariyah scenario. The difference between these scenarios is that in the Stanford simulation, agents start initially neutral and form valence relationships. As a result, the formation of valence levels correlates with personality factors. However, the Hamariyah scenario utilizes more entrenched valence values that are associated with clan structures rather than personal choices. This is evident in the fact that the Hamariyah scenario shows a high correlation between Valence and Ingroup membership that is not present in the Stanford simulation.

Transferability shows a very strong correlation with Ingroup membership in Stanford only. This is due to the fact that certain actions are only available to members of certain groups, another structural factor. This relationship, along with the Hamariyah relationship between Transferability and Conformity, show that this implementation of Transferability depends significantly on the structural factors of the environment.

InGroups and Reference groups also share an interesting correlation. While the Hamariyah scenario shows a strong correlation between ingroup membership and reference groups, indicating that members tend to want to be in the groups they belong to, there is no significant correlation between these in the Stanford study. While in the Stanford scenario, sharing a common ingroup tends to lead to higher valence, the groups were randomly assigned and it appears that some agents might prefer to be in the opposite group. This is consistent with statements from the subjects in the Stanford Prison Experiment (Zimbardo, 2007).

J.1.3 Effect of Collinearity on Attention in Simulation

Due to the significant effects of collinearity, mixed with the number of factors involved, these simulations show that the apparent relative importance of each factor can differ based upon the context. While this was intended to be a simple internal validity test, it provided some much more interesting results. Despite knowing the form of the equation and the correct ground truth values, the effect of covariance between the factors makes their relationship with attention more complicated. The resulting correlations and regressions for each simulation display a situation where certain factors appear more dominant in orienting attention. Additionally, as was shown with the novelty factor, goal oriented agents can demonstrate results that appear counter to controlled experimental results due to feedback dynamics.

These results provide an interesting view of agent based software as a way to look at experiment design. Certain factors displayed a more dominant impact on attention due to situational factors rather than cognitive processes. This analysis of the simulation data provided some interesting insights for empirical verification and also showed good correspondence with findings from other analyses, such as social network science. In this way, this analysis provided a useful internal validity check for some of the higher level emergent properties.

Additionally, this kind of analysis may be beneficial for mocking up empirical studies. While the origin of the data was from a simulation, attempting to examine real-life people interacting encounters similar issues. Especially in empirical studies using many variables, situational effects may result in significant, potentially unexpected interaction between dependent variables. This issue becomes increasingly significant as the complexity of the theoretical model grows. This indicates that in some circumstances, especially for large studies or complex theories, it would be worthwhile to create a simulated analog of the proposed theoretical model and data collection design. This would allow testing to make sure that the experimental design would provide appropriate data to test the model. This approach would also give an indication of higher-level emergent dynamics that would be indicators that the underlying theory might be valid.

J.2 Attention Pearson Coefficient Tables

For an alternative perspective, the Pearson Correlation Coefficients were also calculated to examine these relationships. These tended to show the same trends, but in a few cases hinted at slightly different relationships. Tables J.3 and J.4 show the Pearson Correlation Coefficients for the Hamariyah Randomized and Stanford Hypothesis conditions, respectively. These are the Pearson equivalents of the Kendall analysis examined previously.

Table J.3: Hamariyah Transmission (Random Cond.) Pearson Correlations

	Y	X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Attended (Y)	1.00	-	-	-	-	-	-	-	-	-	-
Novelty (X_0)	-0.02	1.00	-	-	-	-	-	-	-	-	-
Motivation (X_1)	-0.15	-0.02	1.00	-	-	-	-	-	-	-	-
Selection (X_2)	0.02	-0.02	0.04	1.00	-	-	-	-	-	-	-
Authority (X_3)	0.02	-0.02	0.01	-0.00	1.00	-	-	-	-	-	-
Conformity (X_4)	0.14	0.10	-0.26	-0.00	-0.08	1.00	-	-	-	-	-
Similarity (X_5)	0.01	0.00	0.25	0.19	-0.03	-0.03	1.00	-	-	-	-
Transferability (X_6)	0.01	-0.03	-0.18	0.09	-0.05	0.47	0.02	1.00	-	-	-
Valence (X_7)	0.11	-0.14	-0.07	0.14	0.08	-0.01	-0.00	0.01	1.00	-	-
InGroup (X_8)	0.07	-0.08	0.02	0.12	-0.02	-0.02	0.12	0.04	0.40	1.00	-
Ref Group (X_9)	0.08	-0.05	-0.01	0.10	0.01	-0.04	0.10	0.01	0.37	0.82	1.00

One significant difference in the Hamariyah correlations is that the Pearson correlation does not show the moderate connection shown by Kendall (-0.10) between Novelty and Motivated Attention. This indicates that while the values are often ranked in a particular order, their values do not track each other linearly. In the Stanford Analysis, Pearson shows conformity with a smaller negative correlation with attention than in Kendall analysis. Authority and

Table J.4: Stanford Prison Transmission (Hypothesis Cond.) Pearson Correlations

	Y	X_0	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Attended (Y)	1.00	-	-	-	-	-	-	-	-	-	-
Novelty (X_0)	-0.02	1.00	-	-	-	-	-	-	-	-	-
Motivation (X_1)	0.00	-0.03	1.00	-	-	-	-	-	-	-	-
Selection (X_2)	0.17	-0.02	-0.11	1.00	-	-	-	-	-	-	-
Authority (X_3)	0.01	-0.08	-0.20	-0.11	1.00	-	-	-	-	-	-
Conformity (X_4)	0.02	0.03	0.07	-0.14	-0.02	1.00	-	-	-	-	-
Similarity (X_5)	0.15	-0.02	0.03	0.17	-0.01	-0.12	1.00	-	-	-	-
Transferability (X_6)	0.13	-0.09	0.08	-0.02	0.13	-0.00	0.10	1.00	-	-	-
Valence (X_7)	0.23	-0.07	0.04	0.36	-0.10	-0.12	0.56	0.29	1.00	-	-
InGroup (X_8)	0.15	-0.06	0.31	0.01	-0.13	0.10	0.07	0.70	0.36	1.00	-
Ref Group (X_9)	0.02	0.05	0.05	0.05	-0.08	-0.15	0.41	-0.05	0.11	-0.08	1.00

Motivated Attention show the opposite effect, being a much stronger negative correlation, though both are non-negligible. This same change is seen with Authority and Selective Attention. Additionally, the negative correlation between Transferability and Conformity disappears under the Pearson correlation. The Pearson correlation shows a negative correlation between Authority and Valence which does not exist in the Kendall Analysis. The Pearson correlation shows a much weaker negative correlation between InGroup and Novelty.

J.3 Stanford Transmission Quartiles

The full table for the Stanford Quartile examination is presented here, as an additional reference. These tables lists the value for each agent on the metrics for: average first time learned, average number of exposures to learn, average time of first expression, and fraction of actions taken by the agent which use the meme after it is learned. The agent's average value is displayed, as is their quartile ranking. These rankings were used in Section 7.4.2 to generate the quartile classifications. Table J.6 shows the metrics for guards and ThrowInHole, while Table J.5 shows these same metrics for prisoners and Resist. Note that these tables show time as simulation time (steps) rather than the experiment days.

Table J.5: Stanford Resist Meme Quartiles (Prisoners)

	Learning Time		**First Expression**		**Expression Count**		**Exposures To Learn**	
Agent	Step	Quartile	Step	Quartile	% Actions	Quartile	# Exposures	Quartile
S_00	0	0	451	3	0.13	1	0	0
S_01	50	3	52	1	0.37	2	3.77	2
S_02	50	3	551	3	0.002	0	4.10	3
S_03	49	1	61	2	0.24	2	2.50	1
S_04	50	2	51	0	0.43	3	3.50	2
S_05	0	0	48	0	0.043	0	0	0
S_06	49	1	53	2	0.40	2	3.03	1
S_08	50	2	52	1	0.43	3	3.63	2
S_09	50	3	122	2	0.21	1	4.70	3

Table J.6: Stanford ThrowInHole Meme Quartiles (Guards)

	Learning Time		**First Expression**		**Expression Rate**		**Exposures To Learn**	
Agent	Step	Quartile	Step	Quartile	% Actions	Quartile	# Exposures	Quartile
S_11	387	3	399	2	0.032	3	2.27	0
S_12	390	3	413	2	0.018	2	2.50	1
S_13	0	0	66	0	0.015	1	0	0
S_15	95	0	179	1	0.012	1	2.37	1
S_16	239	1	669	3	0.0002	0	3.57	3
S_17	262	2	301	1	0.019	2	3.97	3
S_18	364	3	525	3	0.019	1	3.17	2
S_19	217	1	545	3	0.0008	0	2.70	1
S_20	120	1	122	0	0.032	3	2.73	2
S_21	272	2	280	1	0.026	3	4.67	3

References

Ackoff, R. L. (1971). Towards a system of systems concepts. *Management Science*, *17*(11), 661–671.

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*(1), 3–9.

Adomo, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality.* New York, NY: Harpers and Bros.

Alloy, L. B., Peterson, C., Abramson, L. Y., & Seligman, M. E. P. (1984). Attributional style and the generality of learned helplessness. *Journal of Personality and Social Psychology*, *46*(3), 681–687.

Anderson, D. R. (1981). The effects of TV program comprehensibility on preschool children. *Child Development*, *52*(1), 151–57.

Anderson, J. (1982, September 5). Dance view: Will choreography ever be respected as an art form? *New York Times*.

Asch, S. (1955). Opinions and social pressure. *Scientific American*, *193*(5), 31–35.

Atran, S. (2001). The trouble with memes. *Human Nature*, *12*(4), 351–381.

Atran, S. (2003). Genesis of suicide terrorism. *Science*, *299*(5612), 1534–1539.

Aunger, R. (2002). *The electric meme: A new theory of how we think.* Simon and Schuster.

Axelrod, R. (1973). Schema theory: An information processing model of perception and cognition. *The American Political Science Review*, *67*(4), 1248–1266.

Axelrod, R. (1997). Advancing the art of simulation in the social sciences. *Complexity*, *3*(2), 16–22.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396.

Backer, G., Mertsching, B., & Bollmann, M. (2001). Data and modeldriven gaze control for an activevision system. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*(12), 1415–1429.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: PrenticeHall.

Bandura, A., Ross, D., & Ross, S. (1963). Imitation of filmmediated aggressive models. *Journal of Abnormal and Social Psychology*, *66*(1), 3–11.

Barrowclough, C., & Hooley, J. (2003). Attributions and expressed emotion: A review. *Clinical Psychology Review*, *23*(6), 849–880.

Bell, D. E., Raiffa, H., & Tversky, A. (1988). *Decision making: Descriptive, normative, and prescriptive interactions.* Boston: Cambridge University Press.

Berger, J., & Heath, C. (2007). Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research*, *34*(2), 121–134.

Berman, E., Felter, J., & Shapiro, J. N. (2009). Do working men rebel? Insurgency and unemployment in Iraq and the Philippines. *NBER Working Paper*.

Bharathy, G. K. (2006). *Agent based human behavior modeling: A knowledge engineering based systems methodology for integrating social science frameworks for modeling agents with cognition, personality and culture.* Unpublished doctoral dissertation, University of Pennsylvania.

Bjarneskans, H., Grønnevik, B., & Sandberg, A. (1996). *The lifecycle of memes.* Downloaded from: http://www.aleph.se/Trans/Cultural/Memetics/memecycle.html.

Bjork, R. A. (2003). Interference and forgetting. In J. H. Byrne (Ed.), *Encyclopedia of learning and memory, 2nd ed.* (pp. 268–273). New York, NY: Macmillan Reference.

Blackmore, S. J. (1999). *The meme machine.* New York, NY: Oxford University Press.

Bornstein, R. (1989). Exposure and affect: Overview and metaanalysis of research, 1968-1987. *Psychological Bulletin*, *106*(2), 265–289.

Bradshaw, C. M., Szabadi, E., & Bevan, P. (1976). Behavior of humans in variableinterval schedules of reinforcement. *Journal of the Experimental Analysis of Behavior*, *26*(2), 135–141.

Brown, P., Zavestoski, S., McCormick, S., Mayer, B., MorelloFrosch, R., & Gasior Altman, R. (2004). Embodied health movements: new approaches to social movements in health. *Sociology of Health and Illness*, *26*(1), 50–80.

Bruni, L., & Sugden, R. (2007). The road not taken: how psychology was removed from economics, and how it might be brought back. *The Economic Journal*, *117*(516), 146–173.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction.* Princeton University Press Princeton: RS Foundation.

Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, *64*(3), 363–423.

Canli, T., Zhao, Z., Brewer, J., Gabrieli, J. D. E., & Cahill, L. (2000). Event-related activation in the human amygdala associates with later memory for individual emotional experience. *Journal of Neuroscience*, *20*(19), RC99,1–5.

Carlson, A. B. (1981). *Communication systems, 2nd edition.* McGrawHill.

Carnahan, T., & McFarland, S. (2007). Revisiting the stanford prison experiment: Could participant selfselection have led to the cruelty? *Personality and Social Psychology Bulletin*, *33*(5), 603–614.

Castelfranchi, C. (2001). Towards a cognitive memetics: Sociocognitive mechanisms for memes selection and spreading. *Journal of Memetics–Evolutionary Models of Information Transmission*, *5*.

Chaiken, S., GinerSorolla, R., & Chen, S. (1996). Beyond accuracy: Defense and impression motives in heuristic and systematic information processing. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 553–578).

Chaiken, S., & Trope, Y. (1999). *Dualprocess theories in social psychology.* New York, NY: Guilford Press.

Chaiken, S., Wood, W., & Eagly, A. H. (1996). Principles of persuasion. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 702–742). New York, NY: Guilford Press.

Cherry, C. E. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the Acoustical Society of America*, *25*, 975–979.

Chielens, K., & Heylighen, F. (2005). Operationalization of meme selection criteria: Methodologies to empirically test memetic predictions. In *Proceedings of the joint symposium on socially inspired computing (aisb05)* (pp. 14–20).

Christakis, N., & Fowler, J. (2008). The collective dynamics of smoking in a large social network. *New England Journal of Medicine*, *358*(21), 2249–2258.

Christiansen, M., & Kirby, S. (2003). *Language evolution.* Oxford University Press New York.

Christie, R., & Geis, F. (1970). *Studies in machiavellianism.* Academic Press.

Cialdini, R. B., & Goldstein, N. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, *55*(1), 591–621.

Colwill, R. M., & Rescorla, R. A. (1990). Effect of reinforcer devaluation on discriminative control of instrumental behavior. *Journal of Experimental Psychology: Animal Behavior Processes*, *16*, 40–47.

Combs, A. (1982). Affective education or none at all. *Educational Leadership*, *39*(7), 495–97.

Comrey, A. (2008). The Comrey personality scales. In *The sage handbook of personality theory and assessment: Personality measurement and testing.* Sage Publications Ltd.

Cornwell, J. B., Silverman, B. G., O'Brien, K., & Johns, M. (2002). A demonstration of the pmf-extraction approach: Modeling the effects of sound on crowd behavior. In *11th conference on computer generated forces and behavioral representation* (pp. 107–113).

Costley, C., Das, S., & Brucks, M. (1997). Presentation medium and spontaneous imaging effects on consumer memory. *Journal of Consumer Psychology*, *6*(3), 211–231.

Cowan, N. (2001). The magical number 4 in shortterm memory: A reconsideration of mental storage capacity. *Behavioral and brain sciences*, *24*(01), 87–114.

CrossonTower, C. (1999). *Understanding child abuse and neglect.* Allyn and Bacon.

Cummins, J. (1998). Immersion education for the millennium: What we have learned from 30 years of research on second language immersion. In M. R. Childs & R. M. Bostwick (Eds.), *Learning through two languages: Research and practice. Second Katoh Gakuen international symposium on immersion and bilingual education.* (Vol. 21).

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain.* New York, NY: Penguin Books.

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of Personal and Social Psychology*, *8*(4), 377–83.

Darwin, C. (1902). *Origin of species by means of natural selection, or the preservation of favored races in the struggle for life.* PF Collier.

Dawkins, R. (1976). *The selfish gene.* USA: Oxford University Press.

Delgado, M. R., Labouliere, C. D., & Phelps, E. A. (2006). Fear of losing money? aversive conditioning with secondary reinforcers. *Social Cognitive and Affective Neuroscience*, *1*(3), 250–259.

Dell, G., Chang, F., & Griffin, Z. (1999). Connectionist models of language production: lexical access and grammatical encoding. *Cognitive Science*, *23*(4), 517–542.

Dennett, D. C. (1995). *Darwin's dangerous idea: Evolution and the meanings of life.* New York, NY: Simon & Schuster.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*(1), 80–90.

Dickinson, A. (1985). Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences (19341990)*, *308*(1135), 67–78.

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology.* Teachers College, Columbia University.

Eckhardt, W. (1988). Comment on Ray's "why the f scale predicts racism: A critical review". *Political Psychology*, *9*(4), 681–691.

Edmonds, B., & Moss, S. (2005). From kiss to kids an antisimplisticmodelling approach. *MultiAgent and MultiAgentBased Simulation*, 130–144.

Elison, J. (2005). Shame and guilt: A hundred years of apples and oranges. *New Ideas in Psychology*, *23*(1), 5–32.

Epstein, L. (1999). A definition of uncertainty aversion. *The Review of Economic Studies*, *66*(3), 579–608.

Eysenck, M. W. (1982). *Attention and arousal.* SpringerVerlag.

Fazio, R. H., RoskosEwoldsen, D. R., & Powell, M. C. (1994). Attitudes, perception, and attention. *The hearts eye: Emotional influences in perception and attention*, 197–216.

Finkelstein, R. (2008, May). Defining memes. In *The second symposium on memetics, memory, social networks, and language.* University of Toronto, Toronto, ON.

Fraley, C., & Raftery, A. E. (2003). Enhanced modelbased clustering, density estimation, and discriminant analysis software: Mclust. *Journal of Classification*, *20*(2), 263–286.

Fromm, E. (1973). *The anatomy of human destructiveness.* Holt McDougal.

Gabora, L. (1999). A review of the meme machine by susan blackmore. *The Journal of Artificial Societies and Social Simulation.* Available from http://jasss.soc.surrey.ac.uk/2/2/review2.html

Gaver, W. (1991). Technology affordances. In *Proceedings of the sigchi conference on human factors in computing systems: Reaching through technology* (pp. 79–84).

Gerot, L., & Wignell, P. (1994). *Making sense of functional grammar.* Antipodean Educational Enterprises.

Gibson, E. J., & Pick, A. D. (2000). *An ecological approach to perceptual learning and development.* Oxford University Press, USA.

Gibson, J. J. (1979). *The ecological approach to perception.* Boston, MA: Haughton Mifflin.

Gibson, J. J. (1986). *The ecological approach to visual perception.* Lawrence Erlbaum Associates.

Giddens, A. (1986). *The constitution of society.* Cambridge: Polity Press.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart.* Oxford University Press, USA.

Granger, C. (1969). Investigating causal relations by econometric models and crossspectral methods. *Econometrica*, *37*(3), 424–438.

Grassian, V. (1992). *Moral reasoning: Ethical theory and some contemporary* *moral problems.* Prentice Hall, Inc.

Hafez, H. M. (2006). Dying to be martyrs: The symbolic dimension of suicide terrorism.. In A. Pedahzur (Ed.), *Root causes of suicide terrorism: The globalization of martyrdom* (pp. 54–80). Cass Series on Political Violence.

HaleEvans, R. (2006). Memetics: A systems metabiology. In *Memetics compendium (prepared for darpa memetics workshop)* (Vol. 1, pp. 668–686).

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning.* Hodder Arnold.

Haney, C., Banks, W. C., & Zimbardo, P. G. (1973a). Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology*, *1*, 69–97.

Haney, C., Banks, W. C., & Zimbardo, P. G. (1973b). Study of prisoners and guards in a simulated prison. *Naval Research Reviews*, *9*, 1–17.

Harkins, S. G., & Petty, R. E. (1987). Information utility and the multiple source effect. *Journal of Personality and Social Psychology*, *52*(2), 260–268.

Hastorf, A. H., & Cantril, H. (1954). They saw a game. *Journal of Abnormal and Social Psychology*, *49*, 129–134.

Hayes, B. (2003). *No arms needed, a hero among us.* [Motion picture]. USA: Advanced Medical Productions, Inc.

Heath, R. L., & Bryant, J. (2000). *Human communication theory and research: Concepts, contexts, and challenges.* Lawrence Erlbaum Associates.

Hebb, D. (1949). *The organization of behavior: A neuropsychological theory.* Wiley.

Helsel, D. R., Mueller, D. K., Slack, J. R., & Geological Survey (US). (2006). *Computer program for the kendall family of trend tests.* US Dept. of the Interior, US Geological Survey.

Hermann, M. (2003). Assessing leadership style: A trait analysis. *The Psychological Assessment of Political Leaders*, 178–214.

Herskovitz, M. (1952). *Man and his works.* New York, NY: Alfred Knopf.

Herzberg, F., Mausner, B., & Snyderman, B. (1993). *The motivation to work.* Transaction Publishers.

Heylighen, F. (1998). What makes a meme successful? selection criteria for cultural evolution. In *16th international congress on cybernetics* (pp. 418–423). Namur, Belgium: Association internationale de cybernétique.

Hilmert, C., Kulik, J., & Christenfeld, N. (2006). Positive and negative opinion modeling: The influence of anothers similarity and dissimilarity. *Journal of Personality and Social Psychology*, *90*(3), 440–452.

Hintzman, D., & Block, R. (1971). Repetition and memory: Evidence for a multipletrace hypothesis. *Journal of Experimental Psychology*, *88*(3), 297–306.

HmeloSilver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problembased and inquiry learning: A response to kirschner, sweller, and clark (2006). *Educational Psychologist*, *42*(2), 99–107.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, *117*(4), 500–544.

Holland, J. (1998). *Emergence: from chaos to order* (Vol. 2). Perseus Books.

House, R. (2004). *Culture, leadership, and organizations: The globe study of 62 societies.* Sage.

Howard, D. (1997). Familiar phrases as peripheral persuasion cues. *Journal of Experimental Social Psychology*, *33*(3), 231–243.

Hull, C. (1943). *Principles of behavior: An introduction to behavior theory.* D. Appleton-Century Company, incorporated.

HutchinsonGuest, A. (1989). *Choreographics: A comparison of dance notation systems from the fifteenth century to the present.* Routledge.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology, General*, *110*(3), 306–40.

Jain, S., & Maheswaran, D. (2000). Motivated reasoning: A depth-ofprocessing perspective. *Journal of Consumer Research*, *26*(4), 358–371.

James, W. (1890). *The principles of psychology.* Cambridge, MA: Harvard University Press.

Janis, I., & Mann, L. (1977). *Decision making: A psychological analysis of conflict, choice, and commitment.* Free Press New York.

Johns, M. (2007). *Human behavior modeling within an integrative framework.* Unpublished doctoral dissertation, Citeseer.

Johnston, W. A., & Dark, V. J. (1986). Selective attention. *Annual Reviews in Psychology*, *37*(1), 43–75.

Johnston, W. A., Hawley, K. J., Plewe, S. H., Elliott, J. M. G., & DeWitt, M. J. (1990). Attention capture by novel stimuli. *Journal of Experimental Psychology: General*, *119*(4), 397–411.

Kaelbling, L., Littman, M., & Moore, A. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237–285.

Kahneman, D., & Tversky, A. (2004). Prospect theory: an analysis of decision under risk. *Preference, Belief, and Similarity: Selected Writings*, 549–582.

Kameda, T., Ohtsubo, Y., & Takezawa, M. (1997). Centrality in sociocognitive networks and social influence: An illustration in a group decision-making context. *Journal of personality and social psychology*, *73*(2), 296–309.

Kashima, Y., Klein, O., & Clark, A. E. (2007). Grounding: Sharing information in social interaction. In K. Fiedler (Ed.), *Social communication* (pp. 27–77). Association internationale de cybernétique.

Kelley, G. A. (1955). The psychology of personal constructs. vol. 1. a theory of personality. vol. 2. clinical diagnosis and psychotherapy.

King, M. L. (1998). *The autobiography of Martin Luther King, Jr.* (C. Carson, Ed.). New York, NY: Intellectual Properties Management in association with Warner Books.

Koomey, J. G. (2002). From my perspective-avoiding "the big mistake" in forecasting technology adoption. *Technological Forecasting and Social Change*, *69*(5), 511–518.

Koshland Jr., D. E. (2002). The seven pillars of life. *Science*, *295*, 2215–2216.

Krepinevich Jr, A. (2005). How to win in iraq. *Foreign Affairs*, *84*(5), 87–104.

Kull, K. (2000). Copy versus translate, meme versus sign: development of biological textuality. *European Journal for Semiotic Studies*, *12*(1), 101–120.

Labov, W. (1994). *Principles of linguistic change.* Blackwell.

Lee, D., Itti, L., Koch, C., & Braun, J. (1999). Attention activates winner-take-all competition among visual filters. *Nature Neuroscience*, *2*, 375–381.

Leibenstein, H. (1950). Bandwagon, snob, and veblen effects in the theory of consumers demand. *Quarterly Journal of Economics*, *64*(2), 183–207.

Levine, M., & Shefner, J. (1991). *Fundamentals of sensation and perception.* Brooks/Cole Publishing Company.

Lewin, R. (1997). *Patterns in evolution: The new molecular view.* Scientific American Library.

Lipovetsky, S., & Conklin, M. (2001). Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry*, *17*(4), 319–330.

Lumley, T. (2009). biglm: Bounded memory linear and generalized linear models. *R package version 0.7.* Available from http://CRAN.R-project.org/package=biglm

Lustick, I., & Miodownik, D. (2009). Abstractions, ensembles, and virtualizations: The titration of complexity in agent-based modeling. *Comparative Politics*, *41*(2), 223–244.

Mackintosh, N. (1983). *Conditioning and associative learning.* Oxford University Press.

Mainzer, K. (2007). *Thinking in complexity: The complex dynamics of matter, mind and mankind.* Springer.

Malone, T., & Lepper, M. (1987). Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, learning, and instruction*, *3*, 223–253.

Mantell, D. (1971). The potential for violence in germany. *Journal of Social Issues*, *27*(4), 101–112.

Marder, E., Abbott, L. F., Turrigiano, G. G., Liu, Z., & Golowasch, J. (1996). Memory from the dynamics of intrinsic membrane currents. In *Proceedings of the national academy of sciences* (Vol. 93, pp. 13481–13486). National

Academy of Sciences.

Margolis, S., & Liebowitz, S. (1995). Path-dependence lock-in and history. *Journal of Law Economics and Technology*, *9*, 205–26.

Margolius, B. H. (2001). Permutations with inversions. *Journal of Integer Sequences*, *4*(Article 01.2.4).

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, *50*, 370396.

Matthews, G., & Wells, A. (1999). The cognitive science of attention and emotion. *Handbook of cognition and emotion*, 171–192.

McClelland, D. (1976). *The achievement motive.* Irvington Publishers: distributed by Halsted Press.

Miall, R. C., & Wolpert, D. M. (1996). Forward models for physiological motor control. *Neural Networks*, *9*(8), 1265–1279.

Milgram, S. (2004). Behavioral study of obedience. In N. Scheper-Hughes, S. P. Bourgois, & P. I. Bourgois (Eds.), *Violence in war and peace.* Blackwell Publishers.

Miller, J., & Page, S. (2004). The standing ovation problem. *Complexity*, *9*(5), 8–16.

Mogg, K., Bradley, B., Hyare, H., & Lee, S. (1998). Selective attention to food-related stimuli in hunger: are attentional biases specific to emotional and psychopathological states, or are they also found in normal drive states? *Behaviour Research and Therapy*, *36*(2), 227–237.

Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*, 175–220.

Novani, S., Putro, U. S., & Deguchi, H. (2007). Searching effective polices to prevent bird flu pandemic in bandung city using agent based simulation. In *51st annual meeting of the international society for the systems sciences.* ISSS.

Nye, B. D., Roddy, M., Bharathy, G., & Silverman, B. G. (2007). Monte carlo: Factionsim. In *2007 conference on behavior representation in modeling and simulation (BRIMS).* Norfolk, VA: SISO.

Oakley, T. (2007). Attention and semiotics. *Cognitive Semiotics*, *2007*(15), 25–45.

Oldmeadow, J. A., Platow, M. J., Foddy, M., & Anderson, D. (2003). Self-categorization, status, and social influence. *Social Psychology Quarterly*, *66*(2), 138–152.

Ortony, A., Clore, G., & Collins, A. (1988). *The cognitive structure of emotions.* Cambridge University Press.

Pashler, H. E. (1998). *Attention.* Psychology Press.

Paunonen, S., & Jackson, D. (2000). What is beyond the big five? plenty! *Journal of Personality*, *68*(5), 821–835.

Peirce, C. S. (1931). Collected writings (8 vols.). , *58*.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in experimental social psychology*, *19*, 123–205.

Piaget, J. (1955). *The construction of reality in the child.* Routledge & Kegan Paul.

Piaget, J., Tomlinson, J., & Tomlinson, A. (1929). *The child's conception of the world.* Routledge & Kegan Paul.

Platow, M. J., Haslamb, S. A., Botha, A., Chewa, I., Cuddona, M., Goharpeya, N., et al. (2005). "it's not funny if they're laughing": Self-categorization, social influence, and responses to canned laughter. *Journal of Experimental Social Psychology*, *41*(5), 542–550.

Pohl, R. (2004). *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory.* Psychology Press.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, *109*(2), 160–74.

Premack, D. (1963). *Prediction of the comparative reinforcement values of running and drinking* (Vol. 139).

Prentice, D., & Miller, D. (1996). Pluralistic ignorance and the perpetuation of social norms by unwitting actors. *Advances in Experimental Social Psychology*, *28*, 161–210.

Ray, M. L., & Sawyer, A. G. (1971). Repetition in media models: A laboratory technique. *Journal of Marketing Research*, *8*(1), 20–29.

Ray, M. L., Sawyer, A. G., & Strong, E. C. (1971). Frequency effects revisited. *Journal of Advertising Research*, *11*(1), 14–20.

Reicher, S., & Haslam, S. (2006). Rethinking the psychology of tyranny: The bbc prison study. *British Journal of Social Psychology*, *45*(1), 1–40.

Reiss, S. (2004). Multifaceted nature of intrinsic motivation: The theory of 16 basic desires. *Review of General Psychology*, *8*(3), 179–193.

Rescorla, R. A. (1991). Associative relations in instrumental learning: The eighteenth bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology Section B*, *43*(1), 1–23.

Rescorla, R. A., & Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii* (pp. 64–99).

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*(1), 169–192.

Rogers, E. (1962). *Diffusion of innovations.* Free Press.

Rogers, E. (1995). *Diffusion of innovations.* Free Press.

Rogers, E. (2002). Diffusion of preventive innovations. *Addictive Behaviors*, *27*(6), 989–993.

Roskos-Ewoldsen, D., & Fazio, R. (1992). On the orienting value of attitudes:

Attitude accessibility as a determinant of an object's attraction of visual attention. *Journal of Personality and Social Psychology*, *63*(2), 198–211.

Ruby, J. (1982). *A crack in the mirror: Reflexive perspectives in anthropology.* University of Pennsylvania Press.

Rushton, J., & Irwing, P. (2009). A general factor of personality in the comrey personality scales, the minnesota multiphasic personality inventory-2, and the multicultural personality questionnaire. *Personality and Individual Differences*, *46*(4), 437–442.

Saaty, T. (1996). *Multicriteria decision making: The analytic hierarchy process: Planning, priority setting, resource allocation.* RWS Publications.

Sadoski, M. (1991). A critique of schema theory in reading and a dual coding alternative (commentary). *Reading Research Quarterly*, *26*(4), 463–84.

Schmidt, R. A., & Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis.* Human Kinetics.

Schramm, W. (1963). *The science of human communication.* New York, NY: Basic Books.

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research*, *57*(2), 149–174.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–457.

Shannon, C. E. (1948). A mathematical theory of communication. *Key Papers in the Development of Information Theory*. Available from http://cm.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf

Shibuya, H., & Bundesen, C. (1988). Visual selection from multielement displays: Measuring and modeling effects of exposure duration. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(4), 591–600.

Shoham, Y., & Leyton-Brown, K. (2009). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations.* Cambridge Univ Press.

Sikstrom, S. (1999). Power function forgetting curves as an emergent property of biologically plausible neural network models. *International Journal of Psychology*, *34*(5), 460–464.

Silverman, B. G. (2004). Toward realism in human performance simulation. In J. W. Ness, V. Tepe, & D. R. Ritzer (Eds.), (Vol. 5, pp. 469–498). JAI Press.

Silverman, B. G. (2006). Acasa lab internal working paper. *ACASA Lab*.

Silverman, B. G. (2010). Systems social science: A design inquiry approach for stabilization and reconstruction of social systems. *Intelligent Decision Technologies*, *4*(1), 51–74.

Silverman, B. G., Bharathy, G., Nye, B. D., & Smith, T. E. (2008). Modeling factions for effects based operations, part ii: behavioral game theory.

Computational and Mathematical Organization Theory, *14*(2), 120–155.

Silverman, B. G., & Bharathy, G. K. (2005). Modeling the personality & cognition of leaders. In *Behavior representation in modeling and simulation (BRIMS) conference.*

Silverman, B. G., Bharathy, G. K., Johns, M., Eidelson, R. J., Smith, T. E., & Nye, B. D. (2007). Sociocultural games for training and analysis. *Systems, Man and Cybernetics, Part A, IEEE Transactions on*, *37*(6), 1113–1130.

Silverman, B. G., Bharathy, G. K., Nye, B. D., Kim, G. J., Roddy, M., & Poe, M. (2010). M&s methodologies: A systems approach to the social sciences. In J. A. Sokolowski & C. M. Banks (Eds.), *Modeling and simulation fundamentals: Theoretical underpinnings and practical domains* (p. 227-270). Hoboken, NJ: John Wiley and Sons.

Silverman, B. G., Johns, M., Cornwell, J. B., & O'Brien, K. (2006). Human behavior models for agents in simulators and games: Part i: Enabling science with pmfserv. *Presence: Teleoperators & Virtual Environments*, *15*(2), 139–162.

Silverman, B. G., Might, R., Dubois, R., Shin, H., Johns, M., & Weaver, R. (2001). Toward a human behavior models anthology for synthetic agent development. In *10th conference on computer generated forces and behavioral representation.*

Silverman, B. G., Pietrocola, D., Weyer, N., Weaver, R., Esomar, N., Might, R., et al. (2009). Nonkin village: An embeddable training game generator for learning cultural terrain and sustainable counter-insurgent operations. *Agents for Games and Simulations*, 135–154.

Simon, H. A. (1982). *Models of bounded rationality.* MIT Press.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059–1074.

Sloman, A. (1998). Motives, mechanisms, and emotions. In *Consciousness and emotion in cognitive science: Conceptual and empirical issues* (Vol. 1, pp. 217–233). Garland Publishing.

Snyder, C. R., & Fromkin, H. L. (1980). *Uniqueness.* Plenum Press.

Sparks, J. R., & Areni, C. S. (2008). Style versus substance: Multiple roles of language power in persuasion. *Journal of Applied Social Psychology*, *38*(1), 37–60.

Sternthal, B., Dholakia, R., & Leavitt, C. (1978). The persuasive effect of source credibility: Tests of cognitive response. *Journal of Consumer Research*, *4*(4), 252–260.

Sugrue, L., Corrado, G., & Newsome, W. (2005). Choosing the greater of two goods: neural currencies for valuation and decision making. *Nature Reviews Neuroscience*, *6*, 363–375.

Sutton, R., & Barto, A. (1998). *Reinforcement learning.* MIT Press.

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual Reviews in Psychology*, *33*(1), 1–39.

Tanford, S., & Penrod, S. (1984). Social influence model: A formal integration of research on majority and minority influence processes. *Psychological Bulletin*, *95*(2), 189–225.

Tomasello, M. (1999). *The cultural origins of human cognition.* Harvard University Press.

Townsend, J., & Busemeyer, J. (1995). Dynamic representation of decision-making. In R. F. Port & T. V. Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 101–120).

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136.

Tsuji, A., Ishiko, A., Takasaki, T., & Ikeda, N. (2002). Estimating age of humans based on telomere shortening. *Forensic Science International*, *126*(3), 197–199.

Tweedale, J., Ichalkaranje, N., Sioutis, C., Jarvis, B., Consoli, A., & Phillips-Wren, G. (2007). Innovations in multi-agent systems. *Journal of Network and Computer Applications*, *30*(3), 1089–1115.

Ullmann-Margalit, E. (1990). Revision of norms. *Ethics*, *100*(4), 756–767.

Uttal, W. R. (2001). *The new phrenology: the limits of localizing cognitive processes in the brain.* MIT Press.

Van Gelder, T. (1999). What might cognition be, if not computation? In W. G. Lycan (Ed.), *Mind and cognition: An anthology* (Vol. 38, pp. 1–60). Blackwell Publishers.

Vygotsky, L. S. (1980). *Mind in society.* Cambridge, MA: Harvard University Press.

Weitz, R., & Neal, S. R. (2007). Preventing terrorist best practices from going mass market: A case study of suicide attacks crossing the chasm. In S. S. Costigan & D. Gold (Eds.), *Terrornomics* (pp. 129–144). Ashgate Publishing, Ltd.

West, M., & King, A. (1996). Social learning: Synergy and songbirds. In C. M. Heyes & B. G. Galef (Eds.), *Social learning in animals: The roots of culture* (pp. 155–78).

Whiten, A., Goodall, J., McGrew, W., Nishida, T., Reynolds, V., Sugiyama, Y., et al. (1999). Cultures in chimpanzees. *Nature*, *399*, 682–685.

Wickens, C. D. (1991). Processing resources and attention. In *Multiple-task performance* (pp. 3–34). Taylor and Francis.

Wilkins, J. S. (1998). Whats in a meme? *Journal of Memetics-Evolutionary Models of Information Transmission*, *2*(1).

Windrum, P. (1999). Simulation models of technological innovation. *American Behavioral Scientist*, *42*(10), 1531–1550.

Wood, W. (2000). Attitude change: Persuasion and social influence. *Annual Reviews in Psychology*, *51*(1), 539–570.

Zentall, T. R. (2007). Imitation: definitions, evidence, and mechanisms. *Animal Cognition*, *9*(4), 335–353.

Zimbardo, P. (2007). *The lucifer effect: How good people turn evil.* Rider.

Zukow-Goldring, P., & Arbib, M. (2007). Affordances, effectivities, and assisted imitation: Caregivers and the directing of attention. *Neurocomputing*, *70*(13-15), 2181–2193.