AFRL-RH-WP-TR-2020-0094

HOMO SOCIONETICUS: SCALING THE COGNITIVE FOUNDATIONS OF ONLINE SOCIAL BEHAVIOR

Mark Orr / Ron Fricker
Virginia Polytechnic Inst & State University
300 Turner Street NW, Ste 4200
Blacksburg, VA 24061

Christian Lebiere / Cleotilde Gonzales / David Plaut
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

Peter Pirolli
IHMC
40 S Alcaniz Street
Pensacola, FL 32502

Stephen Reed
USC
3551 Trousdale Parkway
Los Angeles, CA 90007

Monica Capra
Claremont Graduate Univ.
150 E. 10th Street
Claremont, CA 91711

James Moody
Duke Univ.
502 Oregon Street
Durham, NC 27708

Joseph Austerweil
Univ. of Wisconsin - Madison
702 West Johnson Street
Madison, WI 53715

August 2020

FINAL REPORT

Distribution A. Approved for public release; distribution unlimited

AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING
AIRMAN SYSTEMS DIRECTORATE
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE

NOTICE AND SIGNATURE PAGE

AFRL-RH-WP-TR-2020-0094 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FENSTERMACHER.LAURIE.H.1040636281 Digitally signed by FENSTERMACHER.LAURIE.H.1040636281 Date: 2021.01.04 13:21:16 -05'00'

LAURIE FENSTERMACHER
Program Manager
Mission Analytics Branch
Airman Systems Directorate
711th Human Performance Wing
Air Force Research Laboratory

MURDOCK.WILLIAM.P.1048742161 Digitally signed by MURDOCK.WILLIAM.P.1048742161 Date: 2021.01.05 13:24:10 -05'00'

WILLIAM P. MURDOCK, DR-IV, Ph.D.
Mission Analytics Branch
Warfighter Interface Division
Airman Systems Directorate
711th Human Performance Wing
Air Force Research Laboratory

CARTER.LOUISE.ANN.1230249128 Digitally signed by CARTER.LOUISE.ANN.1230249128 Date: 2021.01.06 14:07:51 -05'00'

LOUISE A. CARTER, DR-IV, Ph.D.
Chief, Warfighter Interaction and Readiness Division
Airman Systems Directorate
711th Human Performance Wing
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
8-26-2020

2. REPORT TYPE
Final Technical Report

3. DATES COVERED *(From - To)*
01-10-2017 - 30-6-2019

4. TITLE AND SUBTITLE
Homo SocioNeticus: Scaling the Cognitive Foundations of Online Social Behavior

5a. CONTRACT NUMBER
FA8650-18-C-7826

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
H0V6

6. AUTHOR(S)
Mark Orr - VATech*/UVa (PI) Ron Fricker - VATech
Christian Lebiere – Gleotidle Gonzales – David Plaut - CMU
Peter Pirolli - IHMC
Stephen Read – USC 3551Trousdale Pkwy Los Angeles CA 90007
Monica Capra - Claremont Graduate Univ. 150 E. 10th Str, Claremont CA
James Moody - Duke Univ. 502 Oregon Str, Durham, NC 27708
Joseph Austerweil – Univ. of Wisconsin – Madison, 702 W. Johnson Str., Madison, WI 53715

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Virginia Polytechnic Inst. & State University, 300 Turner Street, NW, Suite 4200, Blacksburg, VA 24061
CMU, 500 Forbes Ave, Pittsburgh, PA 15213
IHMC, 40 S. Alcaniz Str, Pensacola, FL 32502

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Materiel Command
Air Force Research Laboratory
711 Human Performance Wing
Airman Systems Directorate
Warfighter Interaction and Readiness Division
Wright-Patterson AFB, OH 45433

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-RH-WP-TR-2020-0094

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited.

13. SUPPLEMENTARY NOTES
88ABW-2020-0538; Cleared 17 Dec 2020

14. ABSTRACT
The objective of Homo SocioNeticus—a technological platform that was developed by a multi-site, multi-disciplinary team (Virginia Tech, Duke, Carnegie Mellon, IHMC, Stanford, U. Southern California, U. Wisconsin and Claremont Graduate U.) was to provide deep insight into the technical and theoretical best practices for and limitations of accurately scaling models of human cognition, perception, action and motivation to models of populations of cognitive agents that can accurately simulate online social behavioral phenomena (e.g., global information cascades, the evolution of information, and associated effects of social media platform). Central to our approach was the incorporation of substantive social behavioral theory (game theory, social cognition, social decision making, the sociology of human networks) into computational cognitive models of agents (to stand in as individual humans in our simulations). Our team—represented by world class expertise in simulation and computational modeling of populations, cognitive simulation and modeling, computer science, social media analytics, sociology, economics, psychology, and decision theory—developed and operated a technical platform, called the Matrix, to run several large scale use cases of information

15. SUBJECT TERMS
Keywords: Cognitive Science, Cognitive Architectures, Artificial Neural Networks, Bayesian Modeling, Social Science, Neuroscience, Psychology, Sociology, Economics, Complex Systems, Agent-Based Modeling, Multi-Scale Systems.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
U	U	U	SAR	57	Laurie Fenstermacher
					19b. TELEPHONE NUMBER *(Include area code)*

TABLE OF CONTENTS

List of Figures ii

1.0 SUMMARY 1

2.0 INTRODUCTION 2

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES 6

3.1 Overview of Methods 6

3.2 Overview of Cognitive Modeling Approaches 8

3.3 Overview of Social Behavioral Theory 10

3.4 General Operation of *Homo SocioNeticus* Given a Problem/Challenge 12

3.5 Challenge Problem 0 (CP0) 13

3.6 Challenge Problem CP1 (Dynamics of Github, Reddit, & Twitter) 21

3.7 Challenge Problem CP2 29

4.0 RESULTS AND DISCUSSION 36

4.1 CP0 37

4.2 CP1 38

4.3 CP2 41

5.0 CONCLUSIONS 42

6.0 REFERENCES 43

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMNS 49

APPENDIX A - Publications and Presentations 51

LIST OF FIGURES

Figure 1. A Depiction of a well-used Cognitive Architecture, Adaptive Control of Thought—Rational (ACT-R), used to Model Individual-Level Processes 3

Figure 2. The Relation Between Online Social Behavior and Cognitive Models.................. 5

Figure 3. The Two Platforms in Homo SocioNeticus... 8

Figure 4. The General Operation of Homo SocioNeticus given a Problem or Challenge 12

Figure 5. Figure 5. The Organizational Scheme of Homo SocioNeticus in Relation to a Problem Space. 13

Figure 6. The Five-Agent Models Developed for CP0 16

Figure 7. The Apoc Decision Tree for the Baseline Agent in CP0..................................... 17

Figure 8. The Scalability Study Results for Apoc, CP0.. 18

Figure 9. The Social Cognition Model in CP0.. 19

Figure 10. The Full ACT-R Cognitive Architecture with Associated Neurophysiological Under-Pinnings 20

Figure 11. The Base ACT-UP Memory Equations and the Memory Structure used in CP0. 21

Figure 12. Two Approaches I Homo SocioNeticus for Nesting Cognitive and Social Theoretic First Principles in Social Simulation Models 22

Figure 13. Examples of the BDFD Decision Diagram for Agents for Twitter and Reddit in CP1. …………………………………………………………………………………...23

Figure 14. Titration of Cognitive and Social Theory into the BDFD Approach for Twitter .. 24

Figure 15. The Memory Instance in ACT-R for CP1. 25

Figure 16. The User Embedding Method for ACT-R in CP1. 26

Figure 17. Example of a Graph for Computing Familiarity of Context for the Economic Model of Effort for CP1 28

Figure 18. The Formal Intuition Behind the Bayesian Top-Down Modeling Approach Used in CP1. 29

Figure 19. A Schematic of IBLT 30

Figure 20. The Social Roles Extracted for CP2 for Use in ACT-R 31

Figure 21. The Calculation of Preferences for Reactions in CP2 33

Figure 22. The Memory Instance in Act-R for CP2... 34

Figure 23. The Bayesian Top-Down Model Structure for CP2 35

Figure 24. Visualization of the Comparison between Baseline and Cognitive Models for CP0. 37

Figure 25. Comparison of Model Performance from CP0 to CP1.. 38

Figure 26. Comparison of Model Performance from CP1 Challenge Event to Further Developments in the Same Model 39

Figure 27. Comparison of Model Performance from CP1 Challenge Event to Further Developments in the Same Model with Details of Replicates. 40

Figure 28. Comparison of Model Performance in Error in CP2 (yellow [first row] is Baseline BDFD) 41

1.0 SUMMARY

The objective of *Homo SocioNeticus*—a technological platform that was developed by a multi-site, multi-disciplinary team (Virginia Tech, Duke, Carnegie Mellon, Institute for Human Machine Cognition (IHMC), Stanford University, University of Southern California (USC), University Wisconsin and Claremont Graduate University) was to provide deep insight into the technical and theoretical best practices for and limitations of accurately scaling models of human cognition, perception, action and motivation to models of populations of cognitive agents that can accurately simulate online social behavioral phenomena (e.g., global information cascades, the evolution of information, and associated effects of social media platform). Central to our approach was the incorporation of substantive social behavioral theory (game theory, social cognition, social decision making, the sociology of human networks) into computational cognitive models of agents (to stand in as individual humans in our simulations). Our team—represented by world class expertise in simulation and computational modeling of populations, cognitive simulation and modeling, computer science, social media analytics, sociology, economics, psychology, and decision theory—developed and operated a technical platform, called *the Matrix*, to run several large scale use cases of information spread in online social media (e.g., Github, Reddit, Twitter, Telegram).

2.0 INTRODUCTION

The main innovation of the proposed work was a simulation platform suite, *Homo SocioNeticus,* that was designed to provide deep insight into the technical and theoretical best practices for and limitations of accurately scaling models of human cognition, perception, action and motivation to models of populations of cognitive agents that can genuinely simulate online social behavioral phenomena (e.g., global information cascades, the evolution of information, and associated effects of social media platform).

This work was motivated by the fact that there had been little progress in scaling models of human cognition to populations; accuracy measures with respect to population modeling were thus largely unknown; operational insight from such modeling approaches were yet to be tested. At the time of our work, although much technical progress had been made in terms of developing computational and mathematical approaches to online social behavior, it did not incorporate high-fidelity cognitive models—e.g., the explosive growth over the last decade given the theoretical contributions of network science, computer science and engineering and the availability of novel empirical measures of human behavior at-scale (e.g., economic and social media behavior) with fine temporal and person-level resolution. Officially, the field of computational social science was nearly a decade old when we started our project (Lazer, 2010) and still going strong to date (Quattrociocchi, 2017).

The contribution from the cognitive and neural sciences in understanding at-scale social behaviors, writ large, was and still is extremely limited to date, in part because the computational models of individual-level cognition are designed not for scaling to populations, but for detailed theoretical (and sometimes applied) accounts of the dynamics of human cognitive processes, mechanisms, and representations—a computationally heavy enterprise by nature (e.g., **Figure 1**, borrowed from (Taatgen, 2008).

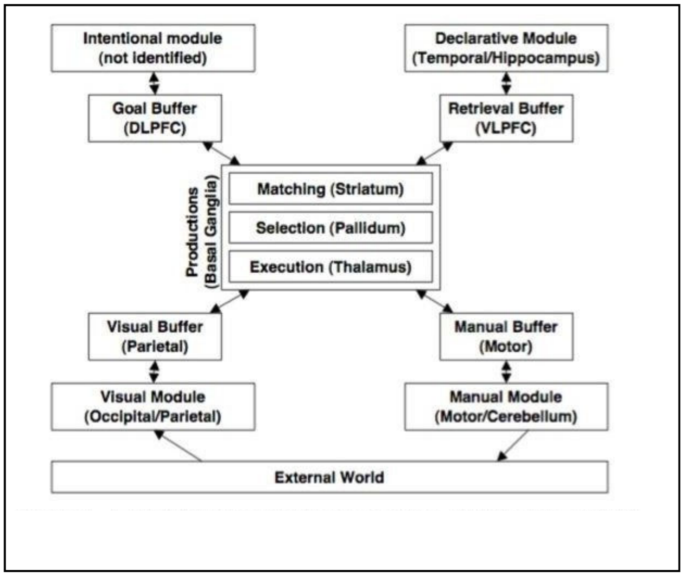


Figure 1. A Depiction of a well-used Cognitive Architecture, Adaptive Control of Thought—Rational (ACT-R), used to Model Individual-Level Processes

Some notable progress in modeling multiple cognitive agents was evident prior to our work. One approach has been to co-opt concepts from cognitive science in relation to cognitive heuristics, e.g., the position of information within a social media interface can drive differences in retweet probability depending on a node's degree (Lerman, 2016). Other approaches were to implement, in social simulations, full-fidelity cognitive models (Orr, 2017; Lebiere et al., 2010; Ritter, 2016), re- implementation of reduced-fidelity cognitive architectures (Reitter & Lebiere, 2010; 2011 Ohlsson, 2010), and use Bayesian formulations of cognitive processes (e.g., Walen, 2017; Burkett, 2010) and language processes (Graf et al., 2016).

The motivation for our work, to implement cognitive modeling into models of social systems, was twofold. First, cognitive mechanisms, when implemented in small social systems, have shown strong interactive effects with respect to social system structure and dynamics—e.g., population-level network assortativity depends on the parameters of the cognitive model (Orr, Zeimer & Chen 2017); memory quality affects social network density (Zhao, 2015); the spread and maintenance of information across a social network of collaborators interacts with properties of the graph (Reitter & Lebiere, 2012); trust dynamics are a function of networks in games of competition and cooperation (Juvina et al., 2015); distributed decision making is dependent on network rewiring probabilities (Romero & Lebiere, 2014); network density and cognitive memory capacity interact in generating socially-based creativity (Bhattacharya & Ohlsson, 2010); and network topology affects the evolution of language (Whalen & Griffiths, 2017).

Second, the micro-level basis of social systems has been well studied in the cognitive modeling community in parallel to studying the interactive effects on small networks—e.g., computational models of attitude formation (Orr, 2013; Orr, 2014; Monroe & Read, 2008; Ehret, 2015); health behavior (see Orr & Chen 2017 for a review); trust and credibility perception and formation (Canini, 2011); experience sampling and person impression formation (Denrell, 2005); belief polarization (Jern, et al 2014); social reasoning and prediction (Baker, Saxe, Tenebaum, 2009; Frank & Goodman, 2012 & 2014). (See Read 2017 edited volume for a survey of current computational approaches in social cognition.) It is of interest that the cognitive modeling approaches to social behavior converge with sociological explanations of network formation and evolution—e.g., homophily, popularity, reciprocity and social balance. A similar relation exists between game theoretic approaches that instantiate Theory of Mind (e.g., in coordination games, Korkmaz, 2016). Game theoretic assumptions have been integrated with concepts drawn directly from cognitive modeling architectures (Gonzalez, 2015).

In sum, our project assumed that the integration of cognitive models into population models of online social behavior would afford potentially unprecedented accuracy in at-scale social simulations because of prior work: 1) smaller scale social simulations using cognitive modeling had produced novel insights into social structures, and 2) the micro-level details of social behaviors that are implicated in online social behaviors were already well studied from a computational cognitive modeling perspective—there were parallel sociological and economic explanations of network-level phenomena that imply similar cognitive mechanisms. The next logical step was to develop new technologies that can push cognitive models at-scale. The fundamental barrier, as we saw it, was in developing technology that could determine which of the many details of task, environment and cognitive processes would be key to the large-scale emergent behaviors we see in populations (Anderson, 2002). This is exactly what *Homo SocioNeticus* was designed to do.

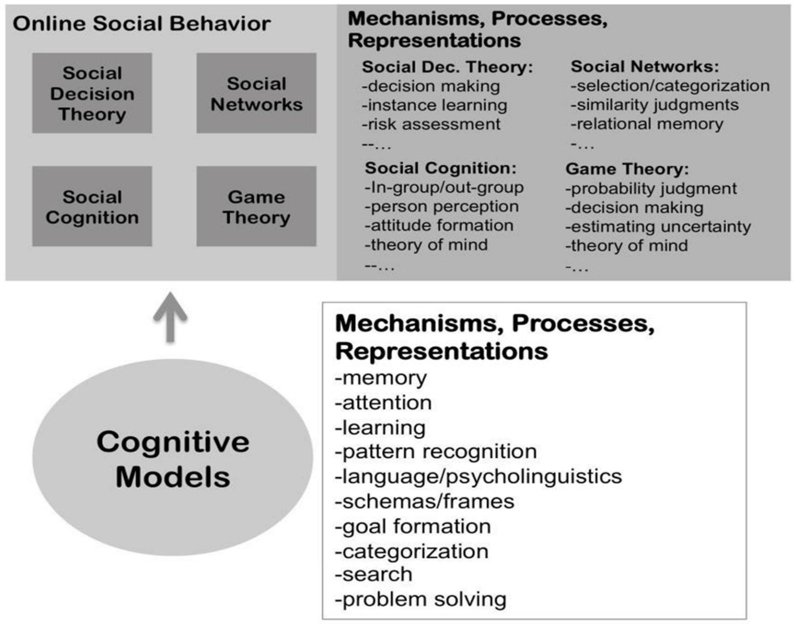


Figure 2. The Relation Between Online Social Behavior and Cognitive Models

The upper half conceptualizes online social behavior as a description of existing social behavioral theory and the implied cognitive mechanisms, processes and representations that are integrated and transformed from basic cognitive models. The lower half shows the basic set of cognitive mechanisms, representations and processes that underlie social behavioral theory. The central concept represented here is that online social behavior implies many basic cognitive processes that have already been formalized by other social disciplines. Homo SocioNeticus leveraged this notion as a socially informed generator of computational cognitive models of online social behavior.

A key conceptual innovation of *Homo SocioNeticus*, represented in **Figure 2**, was the insight that online social behavior falls within a set of related domains of social behavior—game theory, sociology of human networks, social cognition and social decision theory—each of which directly invoke basic cognitive processes. In short, this innovation recognized that the key to building online social simulations accurately and to-scale was the implementation of these domains into cognitive computation formalisms (computational cognitive models) that are directly based in cognitive science and computational psycholinguistics. We thought at the time of starting our work that it is not sufficient to simply build cognitive models from scratch—the social behavioral theory is an essential component of the process.

3.0 METHODS, ASSUMPTIONS, AND PROCEDURES

3.1 Overview of Methods

Our methodological approach provided a technical platform suite for modeling population dynamics of online social behavior at-scale that was grounded in cognitive modeling in its strong form as practiced in the cognitive sciences: *Computational models of human behavior, perception, thought, and action that are constrained by what we know about psychological and neurophysiological processes, mechanisms and representations and that are firmly grounded in and quantitatively evaluated by empirical data (primarily human experiments).* Specifically, *Homo SocioNeticus* leveraged four of the core state of the art computational modeling approaches used in cognitive science: ACT-R, Bayesian modeling, artificial neural networks (ANN), and computational psycholinguistics. Thus, by design, the platform suite was constrained by first principles of human information processing in computational formalisms. We call this platform suite *Homo SocioNeticus* to capture the notion that the advent of social media was a kind of punctuated equilibrium in respect to the human social milieu.

This work was conducted exclusively as a TA1 performer in the Defense Advanced Research Projects Agency (DARPA) SocialSim program. During the reporting period covered by this report, there were three challenges, henceforth called, in order: Challenge Problem 0 (CP0), Challenge Problem 1 (CP1) and Challenge Problem 2 (CP2). Thus, we report the details of our methods in reference to the series of SocialSim challenges. This will afford a clear understanding of the progression of our methods both technically and with respect to social, economic and cognitive theory.

All of the challenges covered in this report had the following general structure. Simulation teams (Technical Area 1 [TA1]) were required to simulate the evolution and diffusion of online information. The basic format of the challenges was to provide TA1s with training data, in the form of event tuples (e.g., a tweet is an event with a user, a parent tweet, a time stamp, and some other features represented in the tuple), from one or more social media platforms (e.g., Twitter, Github, Reddit, Telegram). Given this training data, TA1-trained simulation models were tested at challenge on a hold-out data structure.

The majority of our simulations in the program were agent-based. To this end, we developed a novel agent-based modeling platform for the purposes of the SocialSim program, an open-source simulation platform called the *Matrix Agent-Based Modeling Platform*, shown in **Figure 3-Panel B**. The Matrix was developed with the objective of ingesting modeling efforts in the cognitive science and artificial intelligence communities that were designed to capture some form of human agency in social systems. Thus, it is agnostic to the programming language used to define agents. One can write agent models that work with the Matrix in any programming language, e.g., Python, R, C and more specialized applications coming from the cognitive science and artificial intelligence communities or from formalisms in economics, sociology, social psychology and the decision sciences. Agent model code interacts with the Matrix platform using programming language agnostic interprocess commmunication (IPC) methods (like JavaScript Object Notation [JSON] over transmission Control Protocol/Internet Protocol [TCP/IP]). When using the Matrix platform, model authors can use generic data structures to afford implementation of virtually any social structure (e.g., social interaction graphs) and support the needs of existing cognitive architecture libraries (like ACT-R) and deep neural networks (like Lens, PyTorch, or TensorFlow). Scalability is built into the platform itself such that given a heavy workload (e.g., having large numbers of agents), the computation can be seamlessly distributed across large numbers of machines, without any change in the model code. The *Matrix* codebase is provided as part of the deliverable for this report.

In addition to the *Matrix* platform, we prototyped a platform for top-down (non-agent based) Bayesian social simulation (see **Figure 3-Panel A** for a notional representation of the prototype platform). It functioned, for CP1 and CP2 identically as does the *Matrix* with respect to ingestion of data and production of usable outputs for evaluation. The underlying modeling formalism for this prototype will be used is shown next. (Note that the platform was not hardened as was the *Matrix*, so the codebase provided as part of the deliverable does not operate as a stand-alone platform.)

The top-down platform encodes users interacting through online social media as a *fully* Multivariate Hawkes Process. It is a Point Process, where the conditional probability of a new event given previous events is given by an intensity function:

$$\lambda_{a,e,r}(t) = \mu_{a,e,r} + \sum_{k} \theta_k f_k(t) + \sum_{t_{i,a',e',r'} < t} w_{(a,e,r),(a',e',r')} g(t - t_{i,a',e',r'})$$

for agent a, event type e (e.g., PushEvent or ForkEvent in GitHub), repository r, where $\mu_{a,e,r}$ is their baseline activity, $f_k(t)$ is one of k time-dependent features with weights θ_k, previous events i created by all agents, event types, and repositories, that occur at $t_{i,a',e',r'}$, weights $_{(a,,),(a',e',r')}$ for how related the currently considered event is to previous ones, and kernel $g(t - t_{i,a',e',r'})$. We have developed a novel technique for estimating parameters (w, θ, μ) from training sets. This, in effect, provides a quantitative baseline probability of events based purely on the empirical statistics inherent in the data set. Importantly, our formalism was designed to include information from cognitive and social theory as hierarchical prior distributions on w. For example, consider how people work together in teams on GitHub. There are general schema for how to do so, such as the Fork and Pull scheme. We can formulate this scheme as an expectation as to how weights in this function should look like depending on where the agents are in a current update. Once this is encoded, we can provide a rigorous, quantitative evaluation of how useful that theory is by conducting likelihood ratio tests or Bayes factors of the model with the prior vs. the model that only uses empirical estimates.

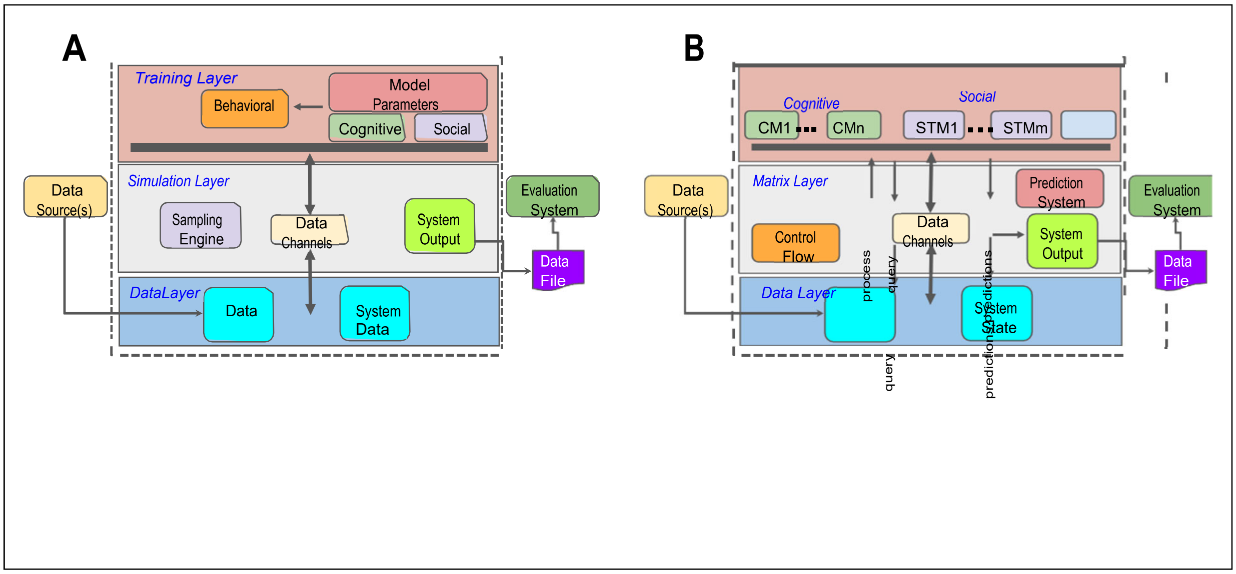


Figure 3. The Two Platforms in Homo SocioNeticus.

In Panel A (left), the proposed top-down Bayesian Simulation Platform is shown in schematic form. Panel B shows the Matrix, Open Source Platform that was developed for agent-based simulation of human social systems.

3.2 Overview of Cognitive Modeling Approaches

ACT-R Approach: ACT-R is a highly modular cognitive architecture, composed of a number of modules (e.g., working memory, procedural and declarative memory, perception and action) that operate asynchronously through capacity-limited buffer interfaces. Each module is in turn composed of a number of independent mechanisms, typically composed of symbolic information processing structures combined with equations that represent specific phenomena and regularities (e.g., power law of practice and forgetting). The architecture includes a number of learning mechanisms to adapt its processing to the structure of the environment. ACT-R hasbeen applied to model human behavior a wide range of applications (see ACT-R web site for over a thousand publications[1]) ranging from simple psychology experiments to decision making to complex dynamic task environments. The combination of powerful mechanisms together with human capacity limitations (e.g., working memory, attention, etc.) provides a principled account of both human information processing capabilities as well as cognitive biases and limitations.

ANNs: ANNs attempt to approximate principles of neural computation in terms of cooperation and competition among simple, neuron-like processing units that learn by adapting their connection weights in response to performance feedback. Initial formulations in the early 1960’s (Rosenblatt, 1962) were subsequently found to have fundamental computational limitations (Minsky & Pappert, 1969). These were overcome in the 1980's by the development of learning procedures, such as back-propagation (Rumelhart et al., 1986) that could train networks with intermediate ("hidden") units between inputs and outputs. Such models led to substantial progress in modeling human behavior across a wide range of perceptual and cognitive domains, although technical and practical limitations limited models to relatively small-scale problems. More recently, the combination of a) the availability of massive amounts of training data from the Internet, b) substantially more powerful computational hardware (exploiting Graphics Processing Units), and c) a number of technical advances, have made it possible to efficiently train networks with a large number of hidden layers on real-world problems, such as object and speech recognition, and this type of "deep learning" (LeCun et al., 2015) is now a state-of-the-art

1 http://act-r.psy.cmu.edu/publication/

machine learning methology.

***Bayesian Approach*:** Bayesian agents assume that their observations d are generated by an unobserved (and noisy) process h. In typical situations, the information provided by the observations is inherently ambiguous, meaning that it could have been produced by more than one hypothesis. For example, did a friend move back in their chair (d), because they are uncomfortable (h_1) or bored (h_2)? A Bayesian agent updates her belief in hypothesis h after observing d, $P(h|d)$, (the posterior distribution) according to the laws of probability. It is proportional to the product of two factors (1) how likely she would be to be given a noisy observation d if h is the true state of the world, $P(d|h)$, and her prior belief in the hypothesis h, $P(h)$: $P(h|d) = P(d|h)P(h)/\Sigma_{h'} P(d|h')P(h')$, where the denominator sums over the two terms for all hypotheses. So, a Bayesian agent would believe her friend is uncomfortable (or is bored) based on how likely she is to move back in their chair when she is uncomfortable (or is bored) and how often in general she is uncomfortable (or bored). Predicting new observations is simple given the posterior belief in each hypothesis: $P(d_{new}|d) = \Sigma_h P(d_{new}|h)P(h|d)$. In short, knowledge is encoded as hypotheses in Bayesian models and background knowledge and constraints on the possible hypotheses is encoded in the prior beliefs. Uncertainty about the state of the world and what the agent would expect to observe if that state were the actual state of the world are encoded as probability distributions. Bayesian modeling is one of the leading formalisms in modern cognitive science for explaining tasks as varied as perception (Kersten, & Yuille, 2003; Ma, 2013) to language understanding (Goodman & Frank, 2016) to social reasoning (Baker, Saxe, & Tenenbaum, 2009) to the acquisition of abstract knowledge and construction of representations (Austerweil & Griffiths, 2013; Tenenbaum, Kemp, Griffiths, & Goodman,2011). One theoretical limitation of Bayesian models is that they require impossibly large number of computations under some conditions--Bayesian agents must sum over all possible hypotheses to make a prediction or decision, which is extremely large for many realistic problems (though see Sanborn, Griffiths, & Navarro, 2010, and Vul, Goodman, Griffiths, & Tenenbaum, 2014).

***Computational Psycholinguistics*:** Human culture is built from the exchange of ideas, accretion of knowledge, and formation of norms. Language is the means by which humans inform, convince, and influence each other, and is hence a crucial building block of culture. Computational psycholinguistics approaches the problem of language use as a quantifiable phenomenon that arises from individual cognition, and studies it with tools from psychology, cognitive modeling, and linguistic theory. In the broadest sense, computational psycholinguistics is practiced any time a precise computational model of a linguistic phenomenon is brought into quantitative contact with human behavioral data. Most relevant for us here, are studies of pragmatics and semantics: how utterances convey meaning in a way that depends on context.

The Rational Speech Act (RSA) framework (see Goodman and Frank, 2016, for a review) has been very successful at explaining language understanding and production for a broad variety of cases. It formalizes, using Bayesian modeling tools, a view of language as social cognition. In essence a listener interprets an utterance by inferring what state of the world would lead an informative listener to make the utterance; this depends on an internal model of a helpful speaker, which itself relies on a model of how a simple literal listener would interpret utterances. Thus, the RSA framework grounds out in a standard semantics, such as a compositional truth-function or a recursive neural net sequence model. But, RSA critically enriches this fixed meaning with context dependence via social reasoning. Using these tools, subtle aspects of language havebeen modeled, including hyperbole and irony (Kao et al., 2014, Kao et al., 2015), vagueness (Lassiter & Goodman, 2015), and implicature (Bergen, Levy, & Goodman, 2016).

3.3 Overview of Social Behavioral Theory

Game Theory: Traditional game theory assumes perfect rationality, rational expectations, and common knowledge of rationality. That is, an agent always behaves optimally or best responds, knows that all other agents know this, and forms beliefs about others' behavior that are consistent with their decisions. Although game theory is a very useful prescriptive mathematical tool, it performs poorly as a descriptive theory of choice. Indeed, experiments with diverse incentives, frames, and subject pools have documented persistent and systematic deviations from the rational prediction (Camerer, 2003). A better approach to describing and predicting strategic choice is to take into account the psychology of decision-making; that is, one should allow for the fact that memory, attention, and computational ability are limited. When considering the psychology of decision-making, one thing becomes clear: choices are the outcome of an imperfect decision process, and decision processes must be explicitly modeled. Modeling the decision-making process requires the explicit modeling of thinking in the form of "reasoning chains: I think that you think that I think…" and implies a solution of a game be derived fromthe point of view of the player rather than the observer. In economics, there have been attempts to explicitly model decision-processes. For instance, Rationalizability (Bernheim, 1984) relies on applying repeated best responses to conjectures about what the players are likely to do; such conjectures, in turn, are best responses. N-level Rationality (Stahl & Wilson, 1993) and Cognitive Hierarchies (Camerer et al., 2004)) allow for heterogeneity of thinking types that differ with respect to their depth of reasoning or how many rounds of best responses players apply. The introspective model introduced by Cabrera, et al. (2004) allows for reasoning chains with error. This introspective process consists of tracing noisy best responses, iteratively, until a stopping rule is satisfied. In general, reasoning chains or introspective models of strategic choice lend themselves well to computational analysis and simulation.

Social Decision Theory: Decision-making processes—e.g., how a human evaluates probabilities, how a human determines value of particular decision alternatives—is another area that lends insights into the problem at hand. In particular, there is a push to advancing current knowledge of how individuals in a group influence each other and how the dynamics of groups help groups and individuals adapt to changing conditions of an environment. Lejarraga, Lejarraga, & Gonzalez (2014) present an initial step towards comparing individual and group adaptation to change. The finding that while groups exhibit best performance during learning compared to individuals, groups are more resistant to change than are individuals. Computational cognitive models and Bayesian approaches suggest that this is due to counterintuitive memory effects: while groups exhibit better common memory than individuals, this better memory results in a form of inertia and inhibits adaptation to change. Individuals in contrast rely on more recent information and given that more recent information is more representative of recent changes, they are able to adapt more successfully to changing conditions. Many questions emerge from these phenomenon: How do individuals in a group influence each other to reach consensus? How do groups with different network structure adapt to change? How do groups of diverse composition adapt to change?

Sociology and Social Networks: Sociological approaches to network structure typically seek to bridge individual behavior tendencies with macro-structural network properties. The fundamental micro processes at work in most affective social networks are homophily, popularity, reciprocity and social balance. Homophily is the simple notion that people prefer to interact with those who are similar to themselves across multiple salient dimensions. In face-to-face networks, the homophilous dimensions have to be balanced against one another due to the availability of relational alters in the setting. Online social networks, on the other hand, tend toward single-dimension interactions where connectivity is driven by preferences on a single dimension (politics, say), and one need not communicate or even be aware of other dimensions. Another key process is preferential attachment which can lead to very unequal distributions of social engagement, particularly for online networks primarily oriented toward information and expertise. However, a third social mechanism, reciprocity, can serve to limit the impact of preferential attachment process. The norm of reciprocity is culturally ubiquitous and is typically the strongest single driver of face-to-face social relations. People have differing tolerances for asymmetry in social relations, but generally people expect affective relations (friendship, love, etc.) to be reciprocated, and they will not continue to pursue "friends" who do not reciprocate. Finally, social balance is the structural analog to homophily; it is well known by aphorisms such as "a friend of a friend is a friend" or "my enemy's enemy is my friend." Social balance serves to coordinate asymmetric ties, as close friends tend to agree on third parties they both admire—and likely influence each other's tolerance of the third's level of reciprocity. When the tendency for reciprocity and social balance are both high, systems will quickly percolate toward dense social clusters limited by people's relational capacity (typically small groups). When balance is high but there is relatively high tolerance of asymmetry in relations, stable hierarchy emerges as friends coordinate on their peers. All of these features/processes interact with each other, such that social balance can magnify the relevance of social features for homophily while reciprocity and homophily can fragment social systems. Recent work related to ecological models for social networks suggests that while each of these primary elements are active in almost all face-to-face relational systems; organizational features constrain the salience of each attribute. The extent to which these sorts of organizational features can affect online social relations is an open question.

Social Cognition: Social cognition addresses a variety of different aspects of social reasoning and behavior, e.g., models of social perception and impression formation explain how impressions of people develop over time (e.g., Ehret, 2015 ; Freeman & Ambady, 2011 : Kunda & Thagard, 1996; Read & Miller, 1993). These theoretical models account for the development of cognitive representations and affective responses (evaluation) to other people, the role of stereotypes and social context on the development of impressions and evaluations, cognitive consistency and coherence among beliefs and the role of motives in biased reasoning (Monroe & Read, 2008; Read & Miller, 1994; Read & Monroe, in press; Thagard, 1989). Further, a variety of models of social decision-making have been developed to include models of reasoned action (Orr et al., 2013; Orr, et al., 2014) model of constraint based decision-making (Gloeckner, 2014) and motivated decision-making. These models capture the role of beliefs and attitudes, motivation, and constraint satisfaction processes in online decision-making, respectively. Another key aspect in social cognition are theoretical models of emotion (e.g., Thagard, 1989), with a particular focus on the role of appraisal processes. Finally, the structure and dynamics of human personality has been modeled in terms of structured, motivational systems. Most of the social cognition literature described above has used artificial neural networks for the past 20 years as theoretical explanations. Recent advances have also leveraged the Bayesian modeling approach. For example, confirmation bias can result from Bayesian agents, given particular assumptions about the environment (Austerweil & Griffiths, 2011). Similar Bayesian mechanisms may drive belief polarization, a phenomenon for which two agents with different viewpoints observe the same evidence and, as a result, diverge even more in their viewpoints (Jern, Chang, & Kemp, 2014). Within the social cognition literature, Bayesian models are a staple computational approach for capturing how people explain, reason, and predict other people's (verbal) behavior (Baker, Saxe, & Tenenbaum, 2009; Frank & Goodman, 2012; Goodman & Frank, 2016), another key facet of online social behavior. The generative capabilities of Bayesian models make them well-suited for modeling online social behavior. Computational psycholinguists offers another perspective on social cognition. Many (perhaps most) interactions between people online are linguistic. Formal models that predict the utterances people will produce to convey an idea (for instance, in email or tweets) and how those utterances will be interpreted, thus form the basic elements out of which complex dynamics can emerge across social networks. For example, recent computational models (Hawkins, Frank, & Goodman, 2017) of classic work on lexical pacts (Clark & Wilkes-Gibbs, 1986) suggest ways of modeling the emergence of new referential forms (which might be hashtags or slang) among tightly interacting groups of people.

3.4 General Operation of *Homo SocioNeticus* Given a Problem/Challenge

The general procedure consisted of three stages in sequence, briefly outlined in **Figure 4.** Stage 1 organized the units and teams to align with a given problem/challenge. Stage 2 was broken up into two parallel and largely independent components. Stage 2A focused on the development of baseline simulation approaches. Stage 2B developed formal cognitive and social theoretic models that were evaluated with respect to capturing aspects of human behavior and social structures inherent in the problem space. Stage 3 integrated the cognitive and social theoretic models into the baseline social simulations. We describe the stages in detail next.

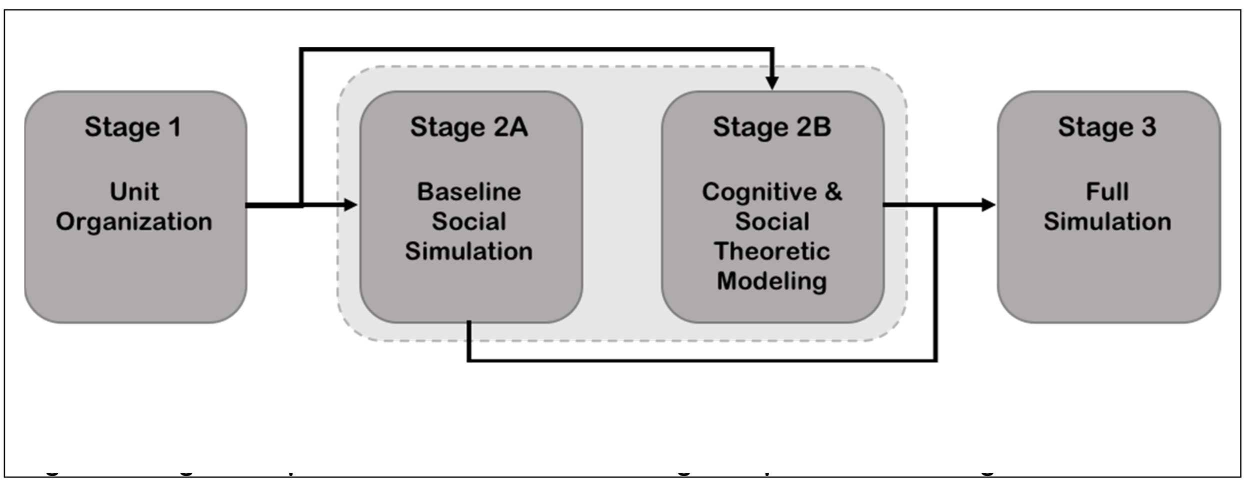


Figure 4. The General Operation of Homo SocioNeticus given a Problem or Challenge.

Stage 1: Homo SocioNeticus Unit Organization: The objective of this stage was to align the efforts of the various units and teams in the larger *Homo SocioNeticus* Organization.

Technical & Scientific Tasking: The *Homo SocioNeticus* organization consisted of expertise from several cognitive modeling and social theoretic approaches. Given a specific problem space (e.g. information spread across multiple online platforms), the first task was to scope out the deliverables and approaches from each unit and teams within units. Figure 5 represents the notion (see the figure caption for details). Teams within each unit coordinated products, efforts, etc. in a principled way towards the objective of building at-scale social simulations.

Baseline Modeling Approach Determination: A central task in this phase was to scope out the details of the baseline modeling approaches and the kinds of cognitive and social theoretic modeling formalisms that were later (Stage 3) integrated into the baseline simulation models. Thus, planning happened jointly and iteratively among all the *Homo SocioNeticus* units. As the *Homo SocioNeticus* organization accumulated experience, it also explored the use of existing tools developed by the organization.

Data Infrastructure Design: The data infrastructure supported all computational modeling and simulation efforts and thus was designed with the following considerations: 1) needs of the baseline simulation approaches, 2) support of development and evaluation of cognitive and social theoretic models. Thus planning, although it did not capture all contingencies, was flexible with respect to needs of the all units. An important part of the data infrastructure design was to determine which units provided definitions of data structures and how this related to the implementation of said data structures (if it was across teams or units, planning for clear communication and evaluation began in this phase).

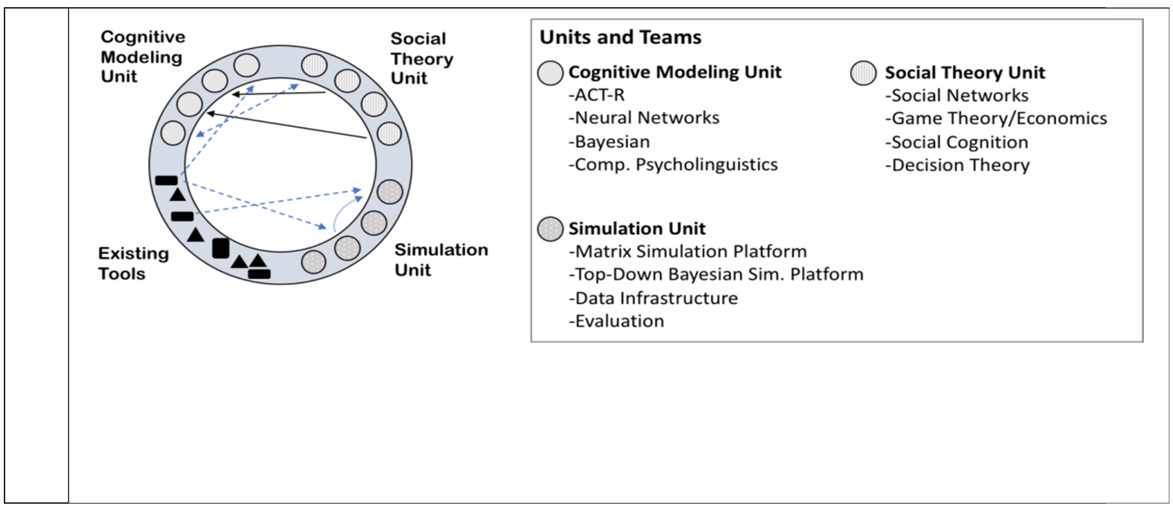


Figure 5. The Organizational Scheme of Homo SocioNeticus in Relation to a Problem Space.

Considered as a dynamic graph, the Homo SocioNeticus organization (a large technical and substantive team) responded to a given problem/challenge by the generation of the appropriate kinds of edges connecting appropriate nodes (prior existing tools were considered nodes on the graph). Some edges transmitted existing code; some provided data constructions; some are bidirectional to illustrate iterative sharing of information or resources.

Stage 2A: Implementation of Baseline Simulations: The objective of this stage was to generate the baseline simulations that served as comparison/controls for simulations that incorporate cognitive processes/representation/theory and social theoretic considerations.

Implementing Baseline Agents/Model Structure Parameters and Training: For the *Matrix* baseline, a principal component was the formalization of the agents (most likely to represent individual- level humans) and the training of the agents. We will provide details of the baseline agent structure for each challenge below; across challenges the baseline agent provided a decision structure that represents action/events at the task level. Further, agent decision structures were specific to an online platform (e.g., Reddit, Twitter, etc.). For problems that involved the simulation of multiple platforms simultaneously, a separate decision structure was implemented for each platform, even when we assumed and simulated the condition that a single individual was active on multiple online platforms (i.e., person X would have a separate decision structure representation on Twitter and Reddit). For the top-down Bayesian baseline, the baseline implementation operated in a different way; it captured the rate at which different event configurations were distributed in the data without explicitly considering a decision structure as in the Matrix. For both the *Matrix* and the top-down Bayesian simulation platforms, the training regimes depended on the parameters of the family of functions used to approximate the data distributions for each approach.

Data Infrastructure**:** The data infrastructure for a problem space was large and complex, capturing event streams of individual users of multiple online social media platforms, data external to online social media platforms, e.g., hourly crypto- currency prices or rare events, textual/verbal behaviors in the form of raw/unstructured text, and naturally, transformations and constructions of said data, e.g., estimation of contact networks from temporal proximity, similarity scores across users or other entities on social media (e.g., among GitHub repository), and latent constructs representing propensity for motivational drives. For analysis and training, the raw data provided for a given problem/challenge was ingested by Apache Spark clusters setup on University of Virginia's (UVA) university's cluster computing system. The training data pipeline used Spark to convert raw data files into Parquet files with well-defined schema. The training data for the different models was then created from this parsed data. During simulation run time, learned model parameters, initialization data, and event data generated during the simulation was stored and processed on locally available embedded data stores (SQLite3, Lightning Memory-Mapped Database [LMDB], and flat comma-separated values [csv] files). During simulation, simulation platforms took over the responsibility of keeping the files synchronized across compute nodes, as new event data was generated. The platforms used embedded datastores instead of databases available over the network (like MySQL, Postgres, Redis etc.) in an effort to minimize compute delays caused by network round trip latency in fetching data during run time.

Stage 2B: Implementation of Cognitive & Social Theoretic Models: The objective for Stage 2B was to provide implementations of the cognitive and social theoretic models for later integration into the full simulation in Stage 3.

Develop Formal Cognitive/Social Theoretic Modeling Formalisms: Working directly with the two simulation platform teams, the cognitive modeling and the social theory units developed formal, computationally implementable models at various levels of abstraction: e.g., provision of dynamic social network contact structure driven by sociological first principles; providing new data features of inputs that have clear theoretical underpinnings from social psychology (motive structures); computation of data structures capturing the processing and semantics of verbal language; the representation of cognitive processes (e.g., memory decay) or representations (e.g., schemas).

Stage 3: Implementation Full Simulation Models: The objective for Stage 3 was to deliver full simulation models that captured cognitive and social theoretic first principles and which may have been embedded or integrated with the baseline simulation models.

Migration of Cognitive & Social Theoretic Models and Data Structures: The data structures used by the cognitive and social theory units were incorporated into the baseline data infrastructure. The nature of the cognitive/social theoretic model determined the method for migration of the models (e.g., a cognitive model that was designed as a modular piece of the baseline decision structure in the *Matrix* baseline simulation model would use existing tie-in points to the baseline code base while a dynamic social network generator would have needed a stand-alone agent).

3.5 Challenge Problem 0 (CP0)

We approached CP0 using the *Matrix* for the development and submission of five agent models, shown in Figure 6. Each of these models was developed by a separate team, but at CP0, only four of the five completed the challenge successfully. The first two in Figure 6 (Baseline and Sociological) did not incorporate cognitive first principles and were considered as *non-cognitive*. We will provide details on the baseline and two cognitive models that successfully completed the challenge. The social media platform we were required to simulate for CP0 was Github.

We conceptualized CP0 and Github, in the *Matrix,* as a user-repo interaction agent-based simulation. Implementation required defining the class of agent (the five agent models described above) that determined agent behavior (where agent is a user), populating a database with a representation of GitHub (the repo and user characteristics in time), and populating the platform with a set of *n* agents. Each of the agents in a simulation were from the same class of agent behavior. Independent of the agent class, all simulations produced the same temporal trace of tuples (user id, repo id, timestamp, and event type) as its main output. (This output was required by the Test and Evaluation [T&E] team.)

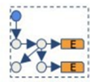
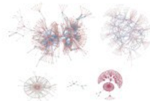
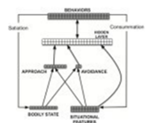
MODEL	AGENT CLASS	DETAILED SPECS.
Baseline (VT)		Causal Bayesian Network based agent. Agent decisions are generated by causal variables that limit the set of choices. Per agent conditional probability tables are used by the model, to learn an agent's propensity to make a choice conditioned upon the causal variables.
Sociological (VT-Duke-Claremont)		Social structure generated from interaction dynamics of training data used to partition system into communities. This structure is incorporated into a regression-based approach for driving individual agent behaviors.
Social Cognition (VT-CMU-USC)		The central notion is that situational context drives behaviors. The key features of the model capture three aspects of motivation: 1) learning from others behavior by observation, 2) learning from direct experience and reward, 3) incorporation of recent past motivational states as part of the current situation.
Cognitive Architecture (CMU-IHMC)		The ACT-R cognitive architecture's long-term memory processes (of past events involving a population of users and associated repos) generates new events by retrieving from memory the most likely user, repo, event type and time interval. Learning processes capture statistical patterns in event distributions. Approximate pattern-matching processes generalize across users and repos.
Bayesian Cognition (Wisc.)	$P(\mathrm{Z},\mathrm{Y}\mid\mathrm{X}) \propto P(\mathrm{X}\mid\mathrm{Z},\mathrm{Y})P(\mathrm{Z})P(\mathrm{Y}).$	Agent behavior as a Hawkes process (a nonhomogeneous Poisson Point Process). The likelihood that a user produces an event where the intensity of the process is a linear combination of time-dependent features. The features are defined in terms of cognitive and social factors (updated over time) and based on cognitively-inspired rules.

Figure 6. The Five-Agent Models Developed for CP0.

Baseline Model: We called the baseline model “Apoc.” Agent decisions were based on a decision directed acyclic graph (DAG) which was trained on the data provided by the T&E. Figure 7 shows the decision structure of Apoc. It was conceptualized as a probabilistic model with conditional independence assumptions. When an agent “has decided” to produce an event, it traverses the DAG to create the event. Ipr, prm, pici, etc. in Figure 7 were parameterized functions that produced the path selection probabilities, and were computed from data.

As this was the simplest agent class for CP0 and the first complex agent we implemented in the *Matrix*, we conducted a scalability study that explored both training time and simulation time across the number of Central Processing Units (CPU) and the number of agents. These results, shown in Figure 8, suggested that, for training, increasing number of CPUs seemed to decrease training time. When considering simulation time, there may have been a tradeoff between number of CPUs and synchronization costs.

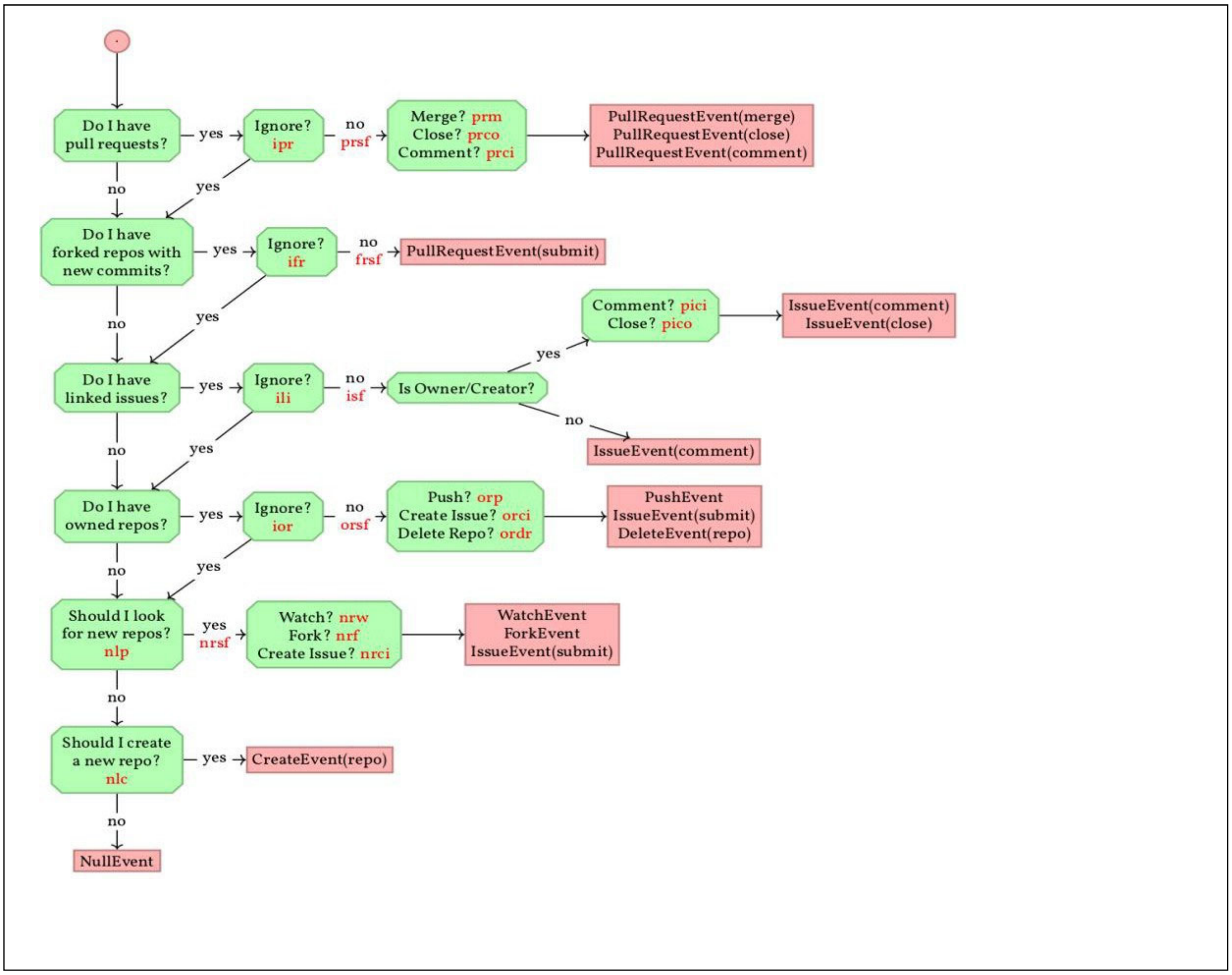


Figure 7. The Apoc Decision Tree for the Baseline Agent in CP0

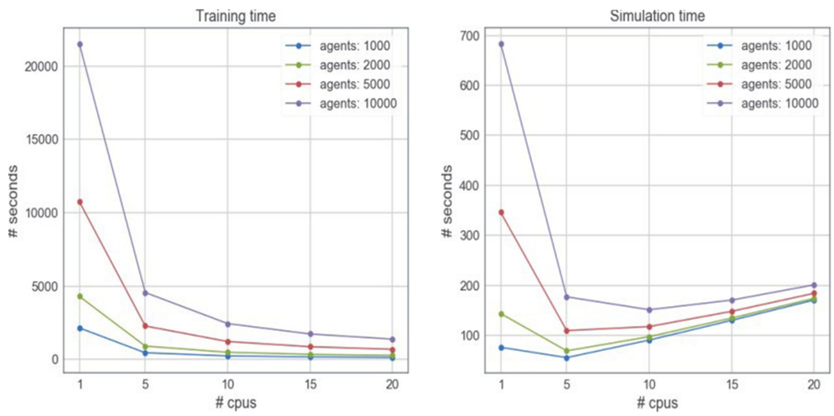


Figure 8. The Scalability Study Results for Apoc, CP0

Social Cognition Model: The social cognition model represented an approach that captured social cognitive aspects of motivation with respect to social behavior on Github. The approach was to designed to capture some of the core aspects of the motivation literature and transform them into a computational model and was motivated by our prior models (Read, 2017) that captured the following characteristics: feedback from actions in the environment via consummation and satiation, approach and avoidance behaviors driven by both features in a situation (food availability) and internal bodily state (e.g., hunger), and individual differences that are represented as differences in sensitivity to reward. These are the theoretical aspects of motivation that drove our development approach.

Figure 9, Panel B, represents the original conception of the model. It was conceived of as a feedforward artificial neural network with a few specialized processes to map onto some important theoretical concepts in motivation theory; the model was conceptualized to represent a single human agent. The input layer was designed to capture features of a Github repo, of the individual user, and individual motives (e.g., attempting to gain social acceptance among a team of software developers). The full input layer fed into a multi-layer hidden unit module with an output layer consisting of the permissible agent actions (e.g., making a pull request) which we call the action layer. The input features of the repo also fed into an autoencoder (a small feedforward network that is trained to reproduce its input). The purpose of the autoencoder was to capture the degree of familiarity an agent had with its current local social context (repos being the context); the error in the autoencoder (as it trains due to exposure to various environments) reflects the inverse of its familiarity which was fed to the action layer. The action layer was designed to have competitive dynamics such that one of n actions would win activation given input. The familiarity autoencoder output gated the action layer such that less error would gate the competitive dynamics less compared to more error. Finally, the objective function was conceptualized to rest on inferred reward—the degree to which an action provided desired reward. Thus, the objective function was conceptualized to served as a kind of reinforcement such that over time the model was supposed to optimize with respect to increasing positive rewards. The original conception of the model, however, was overly complicated given the timeframe for development and, in particular, the difficulty of operationalizing agent motives from the provided data. We provided the development details above to put our final model in its proper theoretical context.

Figure 9, Panel A, represents the model we brought to challenge. It was a significant simplification in its structure yet we attempted to retain some key theoretical components. In effect, we devised an approach to capture feedback from the environment from (e.g., reward sensitivity) the inputs alone. Figure 9, Panel C, shows one such approach.

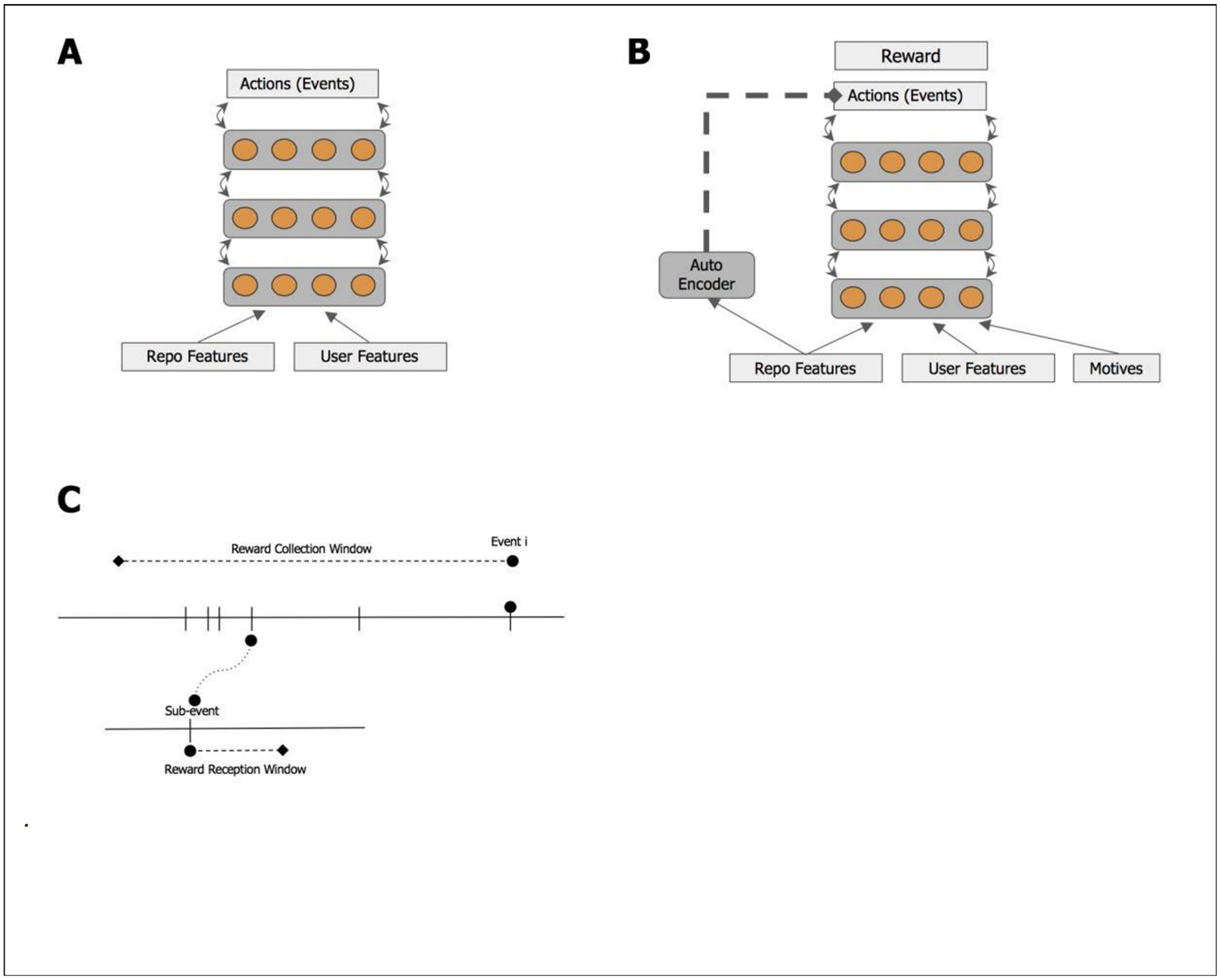


Figure 9. The Social Cognition Model in CP0

Panel A shows the simple version of the model; Panel B shows a more complex version of the model. Panel C shows how we computed reward from past experiences as in input to current social context (a user feature).

Cognitive Architecture Model: The cognitive architecture model, developed in the ACT-R framework (see section above in "Overview of Cognitive Modeling Approaches"), utilized a modular, scaled-down (with respect to computational burden) version of ACT-R called ACT-UP (Reitter & Lebiere, 2010). ACT-UP is a toolkit implementation of ACT-R. It provides an Application Programming Interface (API) for independent cognitive mechanisms and representations (i.e., it does not require the full ACT- R architecture, shown in Figure 10) and thus affords modularization of the implementation of cognitive mechanisms (i.e., this provides access to abstractions of modules of the ACT-R mechanisms as needed for a given task domain). ACT-UP has shown improved time and space scalability by orders of magnitude and has been validated against full ACT-R models.

The theoretical approach was to leverage the scale-free assumptions of cognitive mechanisms from rational analysis of cognitive environments at larger time scales (Newell's rational/social band). In particular, the focus for the Github ACT-UP agent was to represent the history of events of a user as a memory representation for predicting future events. In particular, the cognitive representation of events were characterized as a memory instance (one of many) that captured the history of events/contexts and could be used generatively to predict events going forward (model-tracing). Key characteristics of this approach with respect to the specific ACT-UP implementation were, we hypothesized: 1) base-level activation learning for power laws of recency/frequency, 2) memory activation level and noise may generate cascades / unpredictability, 3) leveraging partial matching for user/repo feature vectors. Figure 11 provides a graphic representation of these characteristics.

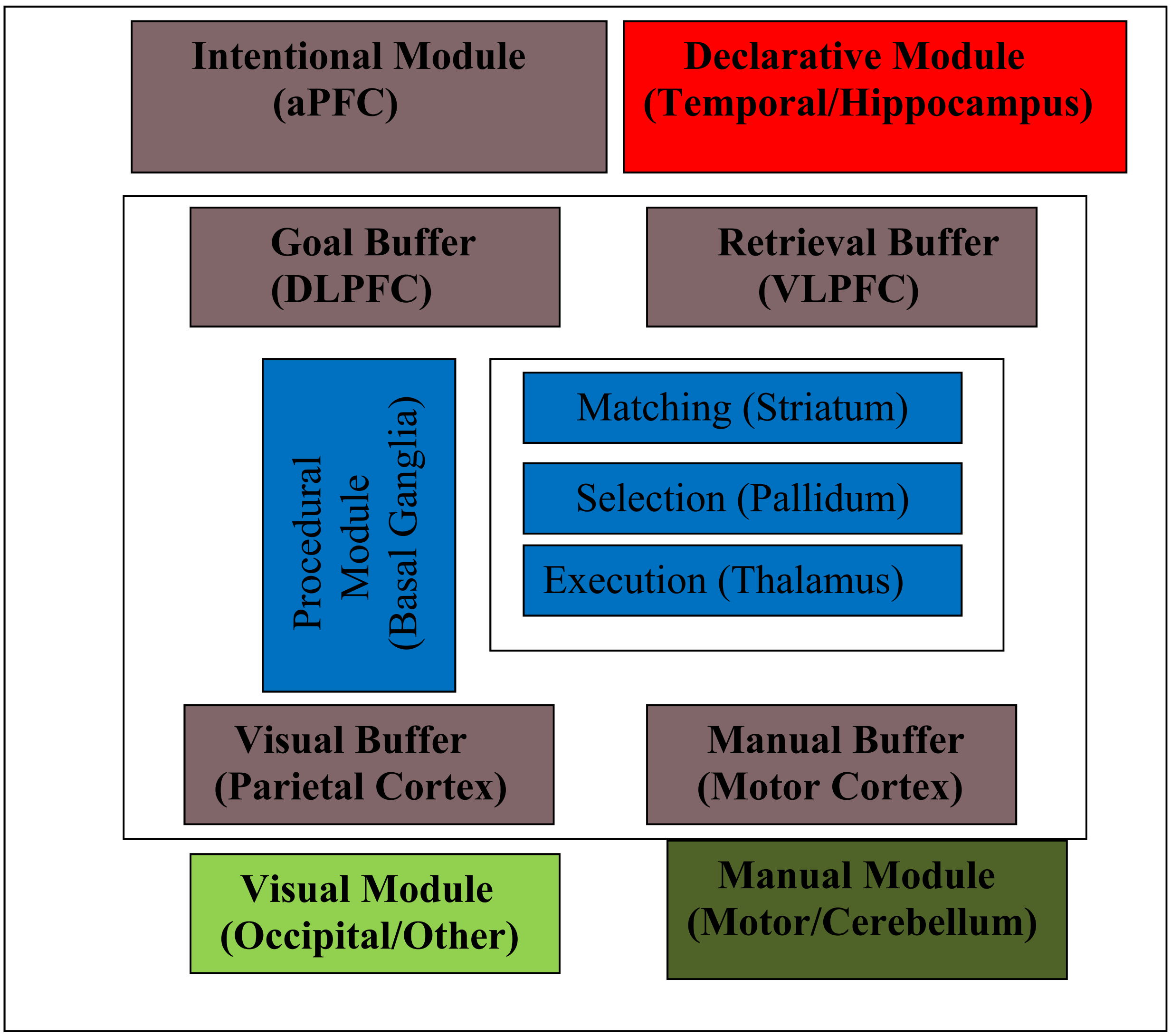


Figure 10. The Full ACT-R Cognitive Architecture with Associated Neurophysiological Underpinnings.

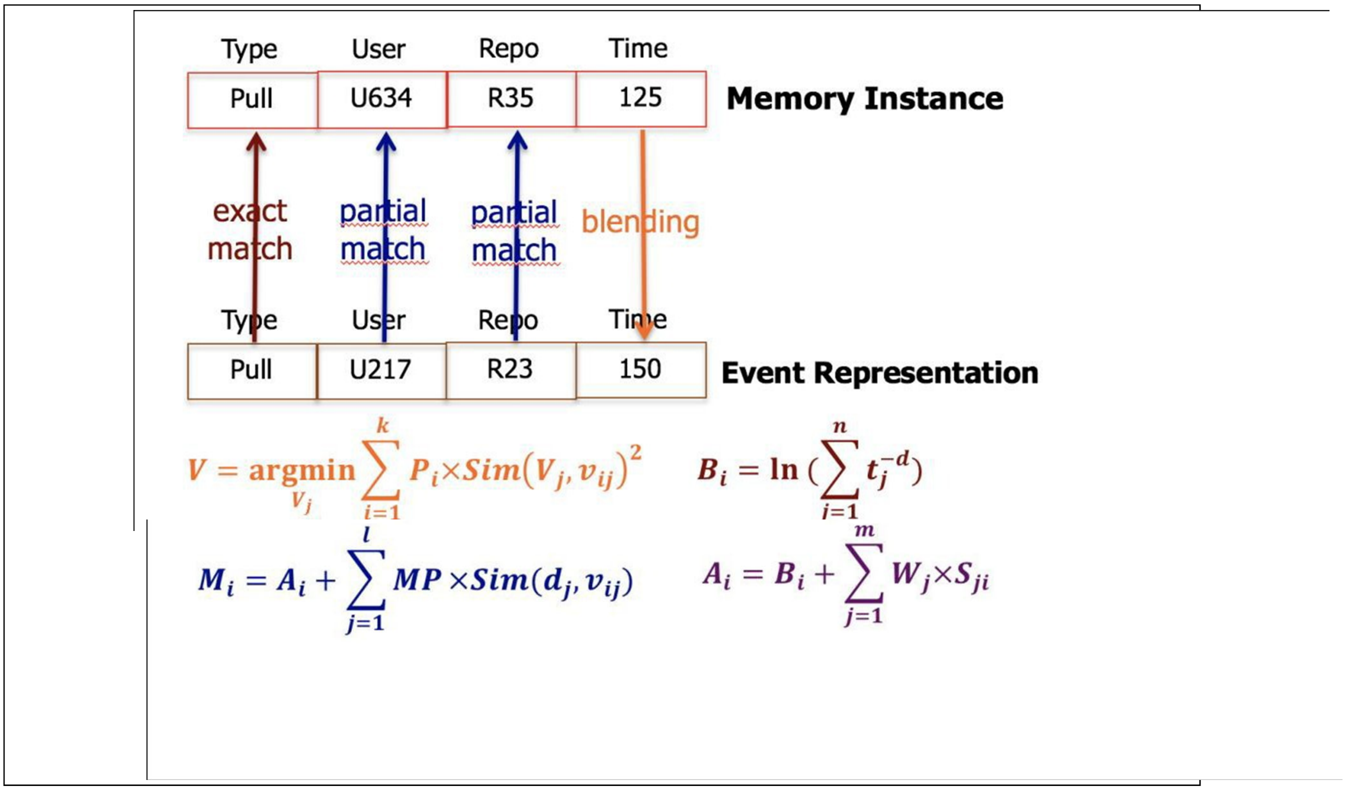


Figure 11. The Base ACT-UP Memory Equations and the Memory Structure used in CP0.
The equations represent the memory processes that underlie the representation of an event for an agent. Each color maps to the diagram arrows.

3.6 Challenge Problem 1 (CP1) (Dynamics of Github, Reddit, & Twitter)

We approached CP1 using both the *Matrix* and the top-down platform (see Section 3.1, Overview of Methods for details). The social media platforms we were required to simulate for CP1 were Github, Reddit and Twitter across two different scenarios (Scenario 1 was the general spread of information online and Scenario 2 was the specific spread by seed) by three different use cases (cryptocurrency, cybersecurity and common vulnerabilities and exposures (CVE). Scenario 2 excluded Github.

A key conceptual innovation of *Homo Socioneticus*, developed for CP1 and shown in Figure 12, was the recognition that the evaluation of the value of our approach will benefit greatly from a systematic, nested comparison within a non-theoretic (cognitive/social) modeling framework or baseline (i.e., a titration of cognitive and social theory into a non-theoretical medium). This recognition implies, further, the implementation of two separate baseline approaches, one to serve as a baseline for cognitive approaches that could be viewed as more top-down in nature (e.g., a Bayesian cognition approach) and another for cognitive approaches that fall more naturally into bottom-up agent-based simulation approaches.

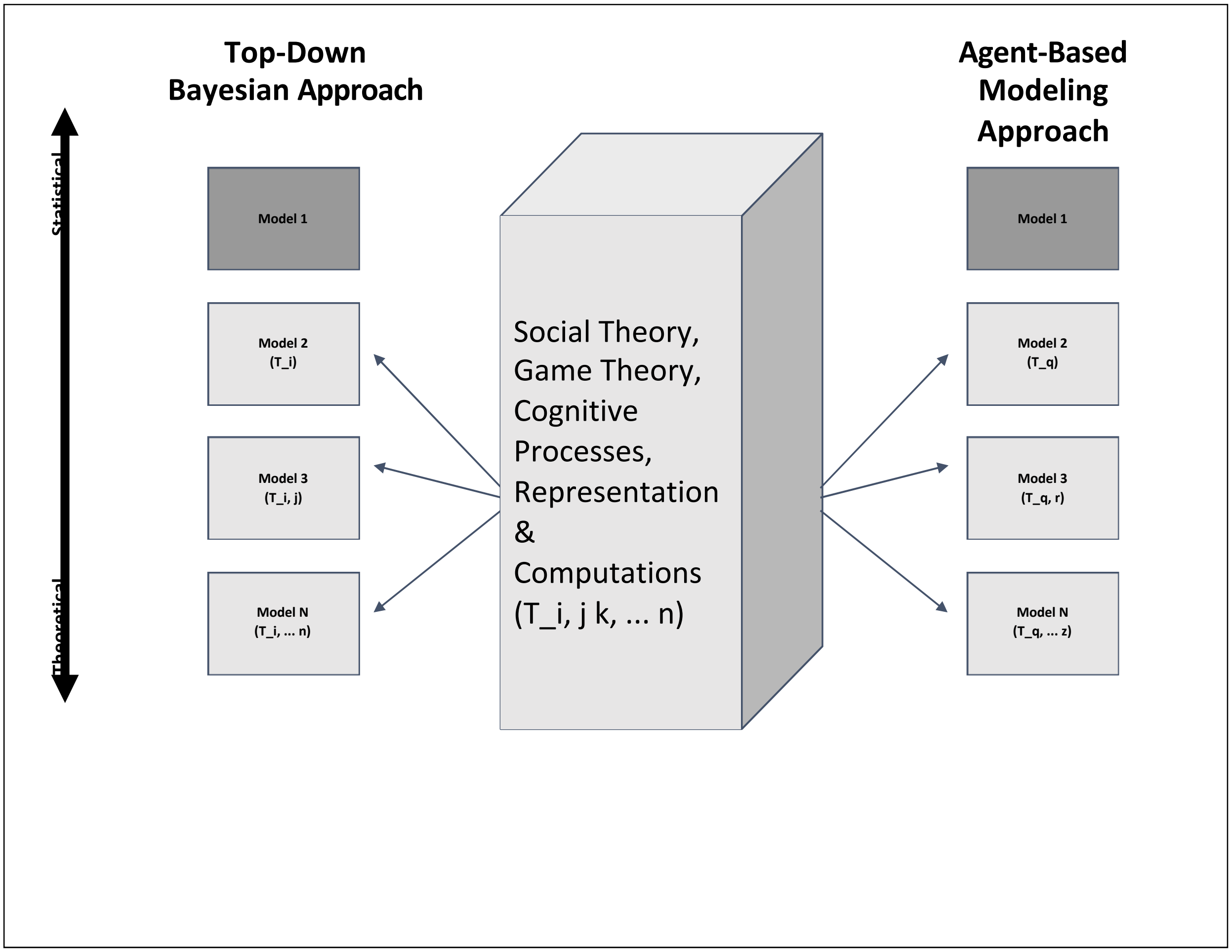


Figure 12. Two Approaches I Homo SocioNeticus for Nesting Cognitive and Social Theoretic First Principles in Social Simulation Models

For each of two simulation approaches, Homo Socioneticus provided a baseline (Model 1; dark grey boxes) that captures the statistics of the data devoid of assumptions from cognitive science or social behavioral theory. The baseline serves as a comparison to models that integrate cognitive and social theoretic first principles to a larger degree (up to Model N).

The baseline modeling effort within the *Matrix* for CP1 using the agent-based modeling approach was grounded in what we call the Bayesian Decision Flow Diagram (BDFD) methodology—an approach to social simulation from a stochastic sampling perspective. This approach, by design, emphasized the modeling of individual-level behaviors and actions (e.g., a single user on a social media platform). By example, imagine a population of human users on the online social media platform Twitter and define an event Ei as a tuple that captures time of the event (τi), the user (ui), the parent tweet (ˆui), and the context (Ci). Formally, then, the baseline model learns a statistical model of the next event a given user ui is going generate given Ci, (C stands for the context of the event that is the set of previous events that have occurred in the system that influences the user ui's decision—e.g., when generating a new tweet, the context that a user draws upon may be tweets that were generated in the past by the user and, potentially other users followed by ui). For purposes of training, an event Ei can be decomposed into its constituent components and affords the estimation/learning of a sequential set of dependencies with respect to the components of the event tuple as shown below:

$$P(\boldsymbol{E}_i \mid u_i, \boldsymbol{C}_i) = P(\tau_i \mid u_i, \boldsymbol{C}_i) \cdot P(\hat{\iota}_i \mid \tau_i, u_i, \boldsymbol{C}_i) \cdot P(a_i \mid \hat{\iota}_i, \tau_i, u_i, \boldsymbol{C}_i)$$

The baseline simulation, given that we've learned P(Ei |ui,Ci) from a corpus of events across a set of users, is computed by sampling from this probability distribution. The sampling is constrained by the assumptions inherent in the *Matrix* (it is a distributed, discrete-time agent-based simulation framework). This latter fact means that we are assuming, during simulation runtime, that agents sample events independently at time *t,* although the distribution captures dependence with respect to prior time periods.

A central feature of this baseline model is that it can directly nest cognitive and social theoretic modeling approaches. It is modular in the sense that it allows for parts of baseline statistical model (the individual conditional distributions) to be replaced with theory guided pieces, allowing us to evaluate the degree to which cognitive/social theory improves or degrades the performance of the model with respect to baseline.

In parallel to how the BDFD was used to titrate theory into the *Matrix* agent-based modeling platform, we used a similar titration approach for the top-down Bayesian social simulation. This method is described above in Section 3.1, Overview of Methods."

Given the titration approach we submitted two classes of model, one was BDFD and the other was the top-down Bayesian approach (we call it the top-down approach henceforth). This is a departure from CP0 in which we submitted five model classes.

Bayesian Decision Flow Diagram Model: The BDFD modeling effort generated several versions, the most developed and important were the baseline and the cognitive implementation. The approach, in the abstract, was identical to the baseline model in CP0 in respect to the production of events (the event payload, naturally, varied over the social media platform in which the agent was embedded). An example of the decision flow for an agent is provided in Figure 13 for Twitter and Reddit.

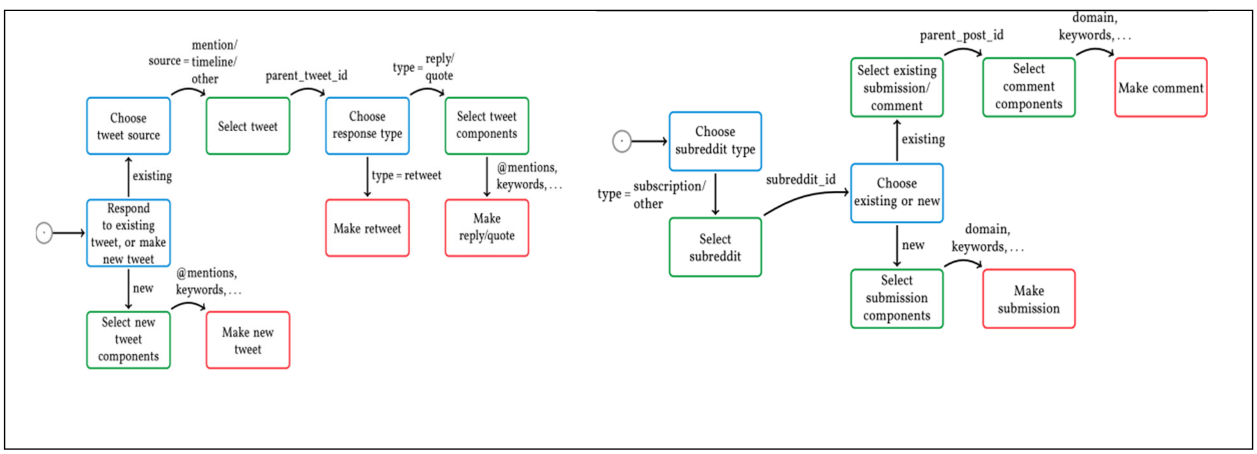


Figure 13. Examples of the BDFD Decision Diagram for Agents for Twitter (left) and Reddit (right) in CP1.

We used the scheme presented in Figure 14 as our model for titration of cognitive and social theory into the BDFD. This kind of titration was only performed for the Twitter agent and for

Scenario 1 in CP1. Using ACT-R/ACT-UP we devised a cognitive model of choice that was swapped out of the BDFD for the “select tweet” module. To represent social theoretic constructs, we also modified the BDFD’s entry point for each discrete time point (Note: forCP1, the *Matrix* platform simulated in discrete time and for each time step selected a fraction of the agents for update; for each update, the BDFD had a model for home many events were produced by the agent, e.g., how many tweets, by the agent) using an economic model of effort. We will review each of these theoretical models next.

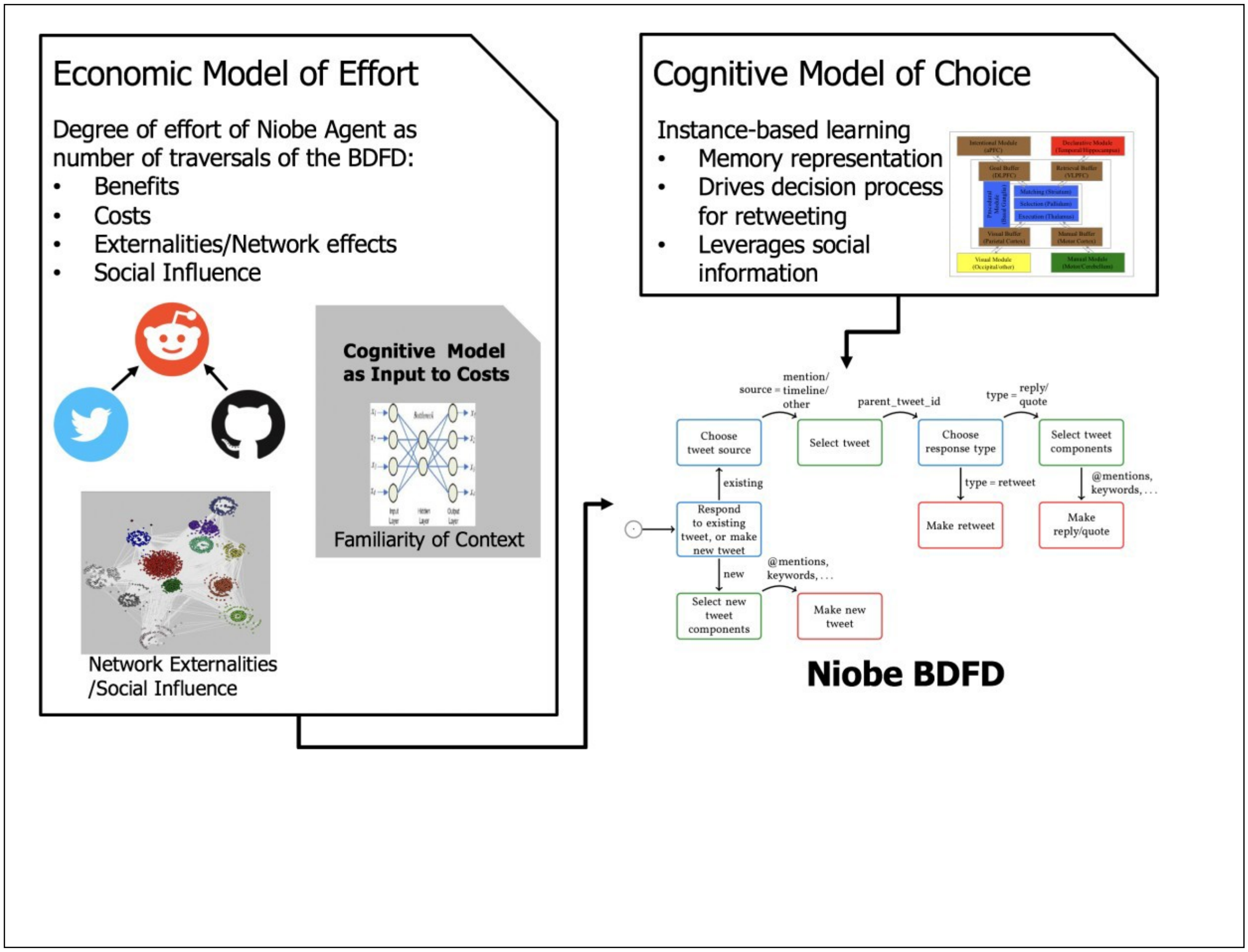


Figure 14. Titration of Cognitive and Social Theory into the BDFD Approach for Twitter Agents in CPI.

See the text for a detailed description of the components. Note: “Niobe” is how we referred to the Twitter Agent.

In the ACT-R cognitive model of choice, the module in the decision graph, as shown in Figure 14, was to select tweet from timeline for reply. For this, the ACT-R model used the Instance-based learning approach whereby the model’s memory holds past decision instances (i.e., past user retweets: positive instances and non-retweets in timeline: negative instances). See Figure 15 for a schematic of the memory instance in ACT-R. The decision process used the blending function in ACT-R to evaluate probability of retweeting to select the tweet with highest probability. This approach was used because it was hypothesized to, potentially, capture the following mechanism in a naturalistic way, i.e., as a natural artifact of the cognitive system: Cascade buildup: power law of practice (e.g., retweet more commonly retweeted users); cascade death: power law decay (e.g., stop retweeting some users over time); Information bubbles: similarity-based matching (e.g., preferential attachment to similar users).

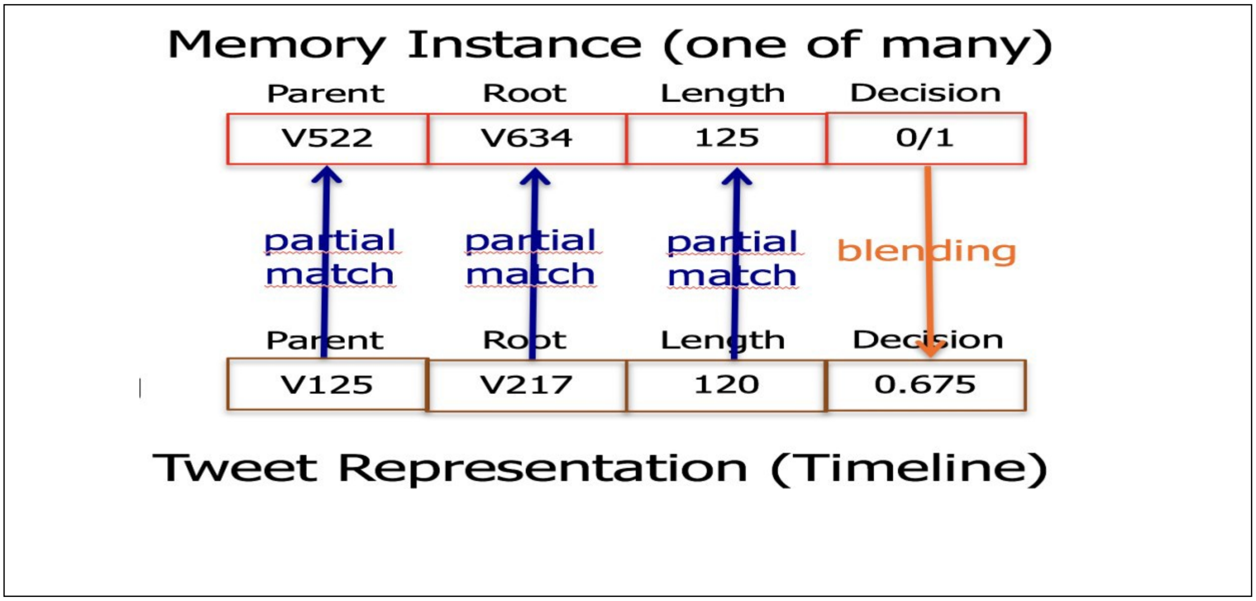


Figure 15. The Memory Instance in ACT-R for CP1.

Another major feature of the ACT-R cognitive model of choice was that it leveraged the concept of similarity by generalization across users (we called it *sparse experience spaces*). Specifically, for the challenge, we generated lower dimensional embedding (LDE) of users based on follower network properties (a 20-dimensional vector with similarity as normalized Euclidean distance). By using K-means clustering over the LDE, we computed clusters as computing nodes for additional generalization (across users). This process is shown graphically in Figure 16.

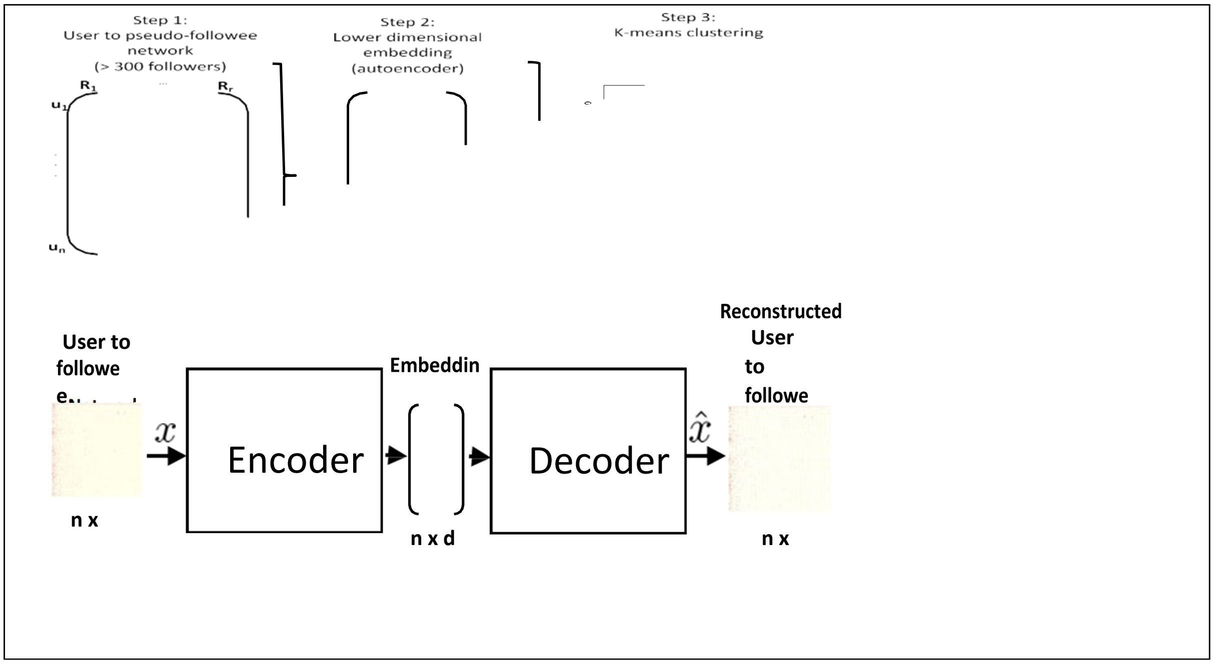


Figure 16. The User Embedding Method for ACT-R in CP1.
See text for details.

For the economic model of effort, which modeled for each agent the number of events generated per simulated time step, we attempted to represent four kinds of theoretically motivated factors: Benefits, costs, externalities/network effects, and social influence. These are described below:

Benefits:

- Just enjoying the activity of tweeting (utility agent gets when he tweets)
- Potential 'likes,' 'retweets,' "positive" replies/quotes the tweet will get-> influence/popularity
- Potential followers the user will get -> influence/popularity

Cost:

- There is some effort needed to create a tweet (which is lower than, say, to retweet)
- This effort is lower, if 'the similarity of a tweet to an agent's past experience' is higher? (if this is a topic s/he is familiar, it is easier to tweet about it)
- Potential, "negative" replies/quotes the tweet will get -> popularity
- Potential followers the user will lose -> popularity

Externalities/Network effects:

- The more others tweet, the more likely the user will tweet: benefit from the number of tweets (probably those s/he can see) on the topic

Social Influence:

- The larger the number of tweeting friends is, the more likely s/he will tweet

We instantiated these theoretical factors as a Poisson distribution model of the daily volume of tweets,

$$Y_{it} \sim Poisson(\lambda_{it})$$
$$\log(\lambda_{it}) = \beta_0 + \beta_1 X_{1it} + \cdots + \beta_n X_{nit} + u_{it} + \varepsilon_{it}$$

where t is time (day) and ii is the user's Twitter ID. β_0 and $\beta_1 \ldots \beta_n$ are fixed effects and u_{it} is between-subject error and ε_{it} is within-subject error. Independent Variables (X_{1it}, ... , X_{nit}) to capture the concepts (cost, benefit, network effects, social influence):

- USER Features: number of followers, number of followees, number of friends, daily tweeting rate
- USER Activity: Posts (tweets, replies, quotes, mentions in the previous day), number of 'favorite's
- Network Effects: Total number of posts on Twitter, Reddit, and Github
- Friend Activity (social influence): Posts by friends
- Topic similarity
- External factors: Closing prices of Cryptocurrency (Bitcoin, Etherium, Monero)
- Control for day of the week

An innovative component of this regression approach was that one of the input variables was computed by a cognitive model of the familiarity of a social media context for an agent. Because we had a computation of user communities (see Figure 17 for an example), we could use this information to estimate a sense of current similarity of current community tweet semantics with the semantics from past experience for a user. The user community graphs capture relations among users that reflect shared interests. For example, in Figure 17 is a network of shared hashtags among users. This graph is weighted in a way that reflects content specificity and is clustered to reflect topical concentration (i.e., groups of people who are interested in the same other people and things).

The cognitive model of familiarity was formalized as an auto-encoder that took the tweets from a user's cluster as input (the trained data reflects the past tweet history of a community; the reconstruction error of any input after training reflects its similarity to the community-based past history). In the regression context, the auto- encoder provided a degree of match between past learned experience and current context and was predictive of the degree of activity for a day given t-1s context. In simulation, this approach was leveraged as a feature of what drives the degree of activity on a given day for an agent. For this procedure we used an in-house vectorization procedure of tweets using blending of crypto-corpus along with already derived token embeddings (at the word-level).

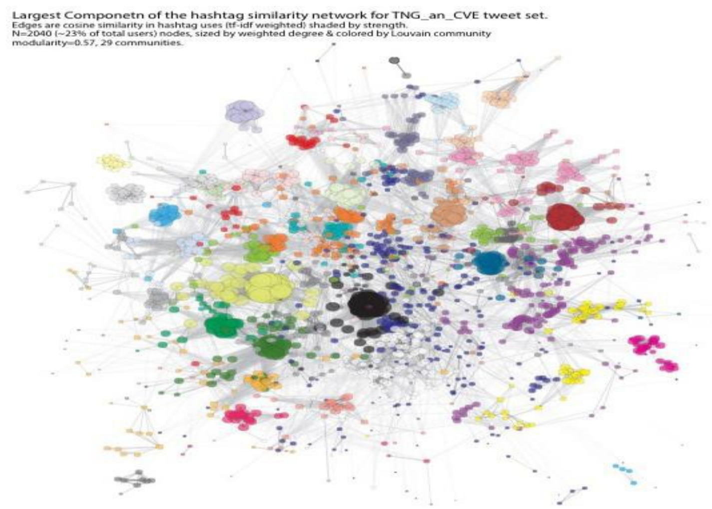


Figure 17. Example of a Graph for Computing Familiarity of Context for the Economic Model of Effort for CP1.

Bayesian Top-Down Model: As described above in "*Overview of* Methods," the top-down platform encoded users interacting through online social media as a *fully* Multivariate Hawkes Process. This approach for CP1 was only developed for Github (Scenario 1). We describe this in some detail because it was not described for CP0 (the model did not make it through the challenge process for CP0). Using this approach we conceptualized the challenge problem in the following way: A GitHub repository was considered a **real-world product** of collaborative cognition. Individuals spontaneously organize to solve a common problem together. Our approach was to develop a model of this process from first principles, where the **repository** was the mind/agent which encapsulates the idea of collaborative (e.g., Searle, 1995) and extended cognition (e.g., Clark & Chalmers, 1998) as stochastic processes. This approach was developed because we posited that it had promise for answering difficult scientific questions and develop novel machine learning methods. Further, this approach afforded the evaluation of interventions and counterfactuals such as, what would have happened to cryptocurrency prices over last few months if event-stream did not get exploited? Finally, the multivariate Hawkes process could model interactions between GitHub events, Cryptocurrency prices, and other social media externalities and captures: 1) The bursty nature and structured relations between events built into model; 2) it can examine importance of theories by nested model comparisons; 3) it can focus on particular past and possible future events and examine how changes would affect currency prices, Twitter feeds, etc.; and, 4) it is interpretable while retaining scale and power and quantitative testability. Figure 18 attempts to characterize and given intuition to some of the formal properties of the modeling approach.

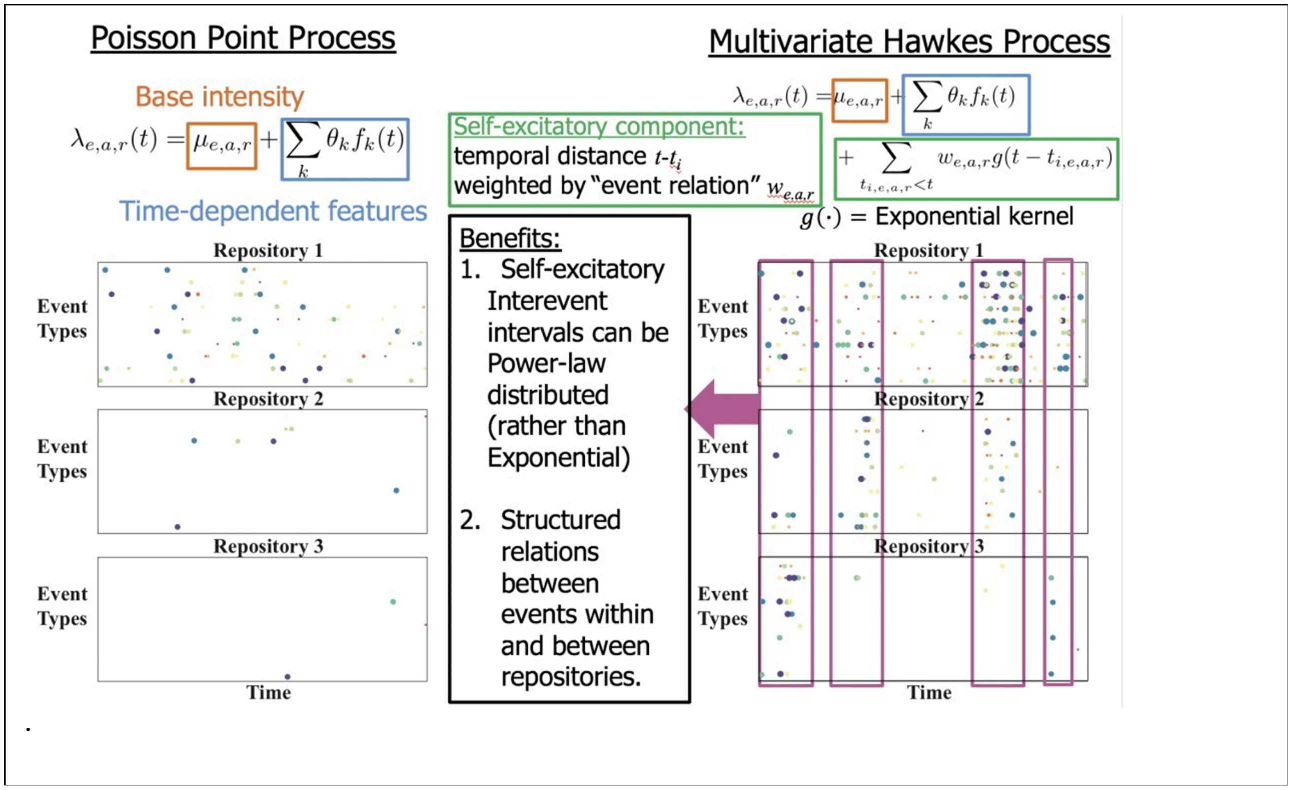


Figure 18. The Formal Intuition Behind the Bayesian Top-Down Modeling Approach Used in CP1.

3.7 Challenge Problem 2 (CP2)

We approached CP2 using both the *Matrix* and the top-down platform (see "Overview of Methods" for details) as we did for CP1. The social media platforms we were required to simulate for CP2 were Github, Reddit and Twitter and Telegram across two different scenarios by three different use cases (cryptocurrency, cybersecurity and CVE). Scenario 1 excluded Telegram; scenario 2 excluded Github and Reddit.

Bayesian Decision Flow Diagram Model: The primary objective for CP2 and the BDFD was to move beyond the method we employed for CP1—the insertion of social, economic and cognitive theory at select nodes to prototype the BDFD approach. For CP2, we explored a more cognitive deepening of each node grounding it in instance-based learning theory (IBLT). The social and economic theory fed into the IBLT at the node level of the BDFD.

The objective, specifically for the BDFD for CP2 was to move towards full integration with respect to the following:

Cognitive Science:

- Decision making apparatus
- Instance-based memory representation
- Learning
- Implementation in ACT-R

Economics

- Expected utility
- Details of functional form and theoretical inputs/functions/constructs
- Social roles & clusters (also Sociological)

Sociology

- Social roles
- Social clusters

The approach for doing so was to represent the economic and sociological theory as a cognitive memory-based mechanism of decision making implemented exclusively within in ACT-R. The theoretical basis for this implementation was taken directly from Instance-Based Learning (IBL) theory (shown schematically in Figure 19).

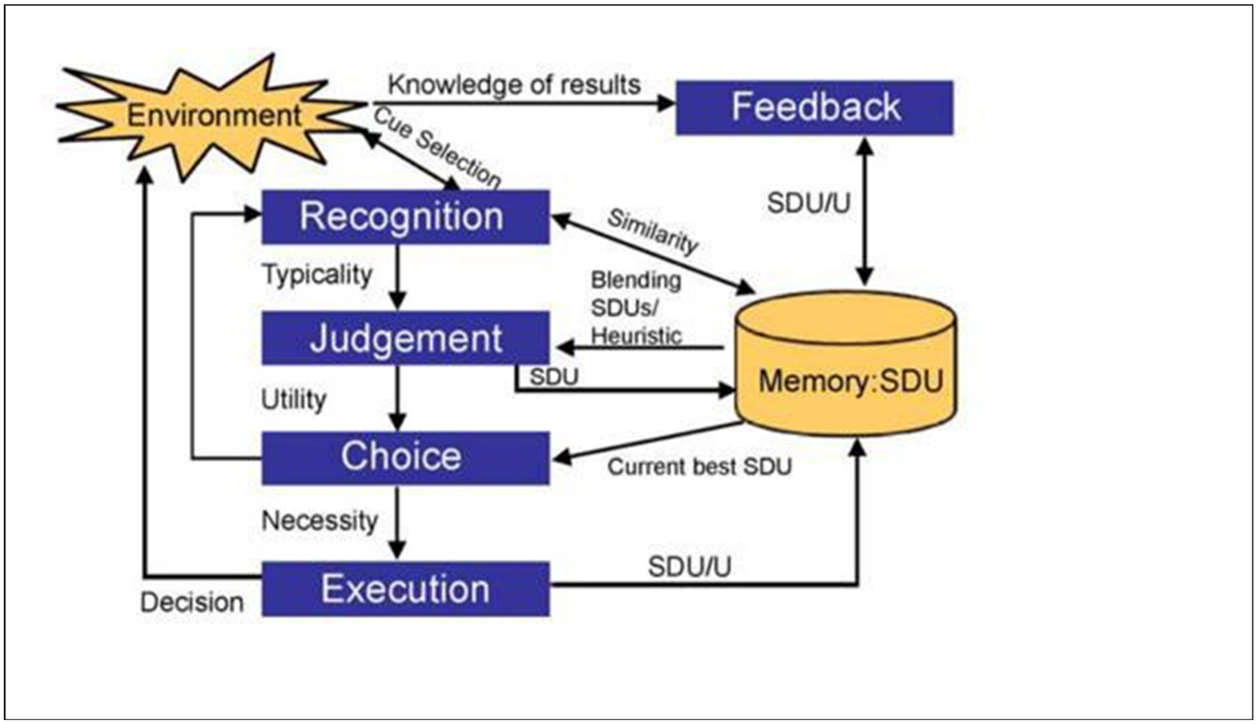


Figure 19. A Schematic of IBLT.

Next we review the technical details of each component: social role generation (sociology), utility (economic) and the integration of these into the ACT-R IBLT modeling framework.

The sociological social role generation approach modeled the relational logic of twitter users by building shared activity (term use, reacting to similar messages) and interaction (replying to or mentioning each other networks, then identified common roles across these networks). This was motived by classic network ideas of White, Boorman & Breiger (1976): *actions embody informal rules and norms, so equivalent actors should behave similarly***.** The agents were assigned to one of 8 substantive roles based on contact patterns across constituent subnetworks. Figure 20 shows an example of this using the Twitter training data for CP2 (it is a sub community of the largest community within the pooled all-relations summary network. N=1046 nodes. Layout uses the Fruchterman Rheingold algorithm, with post-layout adjustments for overlap and pendant (degree=1) placement. Colors track roles. Line thickness and shade capture the strength of the relation).

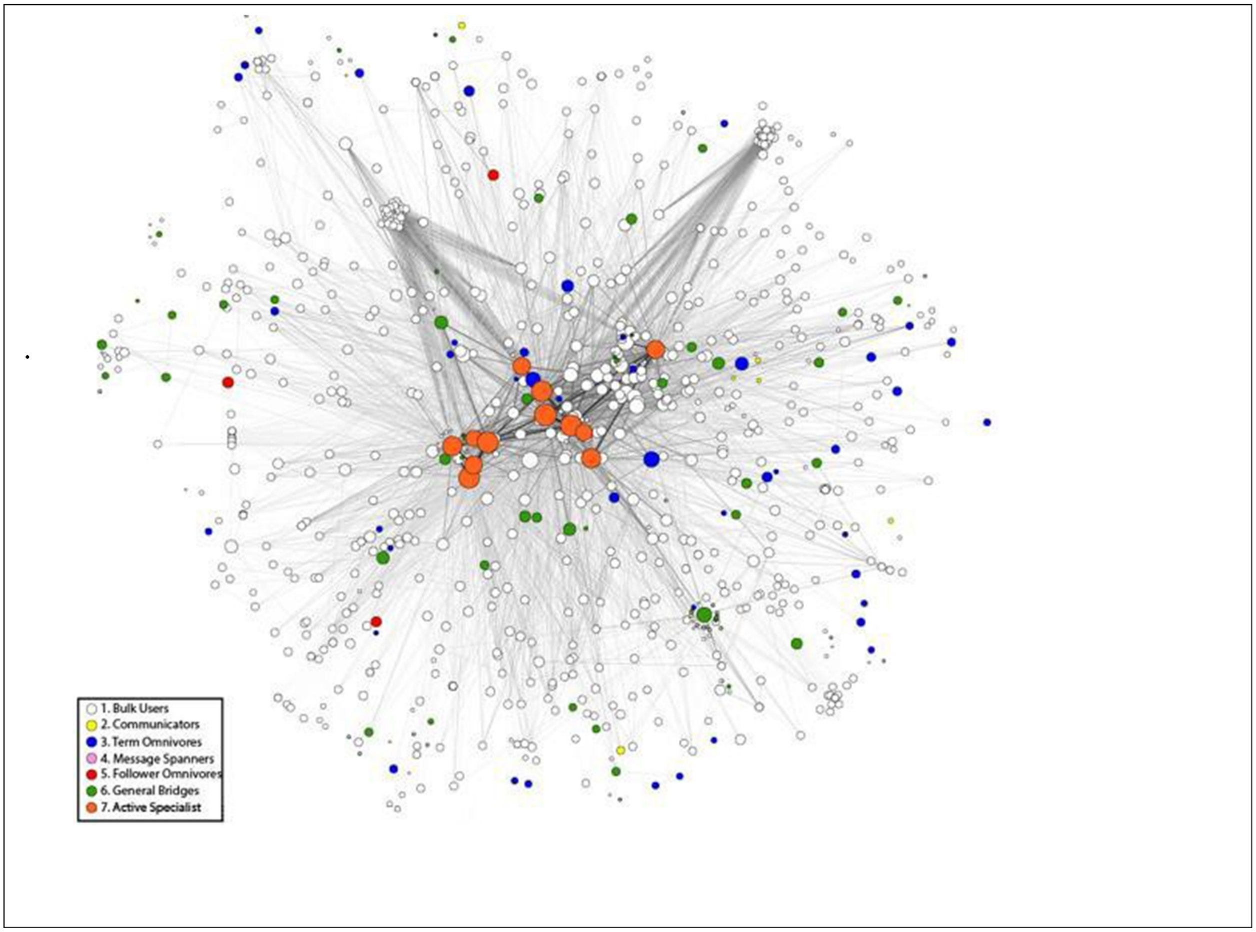


Figure 20. The Social Roles Extracted for CP2 for Use in ACT-R.

The economic utility functions were computed for each role. The idea was as follows.

- Use of ***roles*** to define types based on their degree, weighted degree, and bridging scores across component networks plus the (normalized) sum network.
- The general theory here was that people in the same role will act the same way.
- We estimated the utility function parameters with ***type-specific coefficients.***

Specifically, we assumed that: 1) the agents' objective was to maximize the interaction/engagement (retweet, reply, quote) to their posts, 2) agents differed in their preferences to these reactions, and 3) agents acted optimally, and use their tweet features to estimate their preferences to the interactions. Formally, the model was,

$$U_t(\text{retweet, reply, quote}) = f\,(\text{tweet features, user features})$$

where t represents the identified rols:

Role 1: lower degree across all relations than average and much lower (usually 0) bridging scores.

- Role 2: bridges on the "direct" communication networks (reply & mention combined, but here mainly reply).
- Role 3: bridges in the co-hashtag network but not the others; interested in a diverse set of

#topics.

- Role 4: bridges in the retweet (message) network.
- Role 5: bridges in the retweet (user) network.
- Role 6: high degree and high bridging across lots of different types of networks; very active.
- Role 7: high degree, but low to zero bridging; active in a very limited domain.
- Roles 8 and 9 are outliers. ***very*** high degree (the one person in role 9 is connected to 25K people in the direct follower network!); possibly bots.

To compute this scheme with the given Twitter data, for each individual type (roles 1—8), 3 models were developed for each reaction type (retweet, replies, quotes) and each tweet was modeled as:

$$\log(Y_i) = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_n X_{ni} + \varepsilon_i$$

where Y is a binary variable, $YY = 0$ if the tweet did not receive any reaction, $YY = 0$ if it received at least one reaction. The independent variables ($X_{1i}, \ldots, X_{ni}$) captured tweet features (e.g., is a retweet, reply, quote, url), global features (activity, coin activity), and user features (number of tweets and retweets that were shared, number friends, and followers). Then, the three models were combined to compute the expected reaction type distribution for each role. Figure 21 shows this schematically.

The cognitive ACT-R/ACT-UP model implemented the sociological and economic utility functions as agents in the *Matrix.* Specifically, this was a cognitive model of action choice for which the model had to choose between tweet, retweet, reply and quote. This was considered in one step or two (new/old content). Further, it represented an instance-based learning approach where the model memory held past decision instances (i.e., associated decision context, action chosen, resulting outcome(s) and outcome(s) utility). An iterative decision process used iterative blending to evaluate expected utility of each action and to then select the action with highest expected utility as shown in Figure 22. The model assumed that due to bounded rationality and cognitive biases it would exhibit the following:

- Recency bias--most recent experiences would have disproportionate effect
- Anchoring bias--most common experiences would persist after relevance lost
- Sampling bias--initial experience would lead to lasting risk aversion.

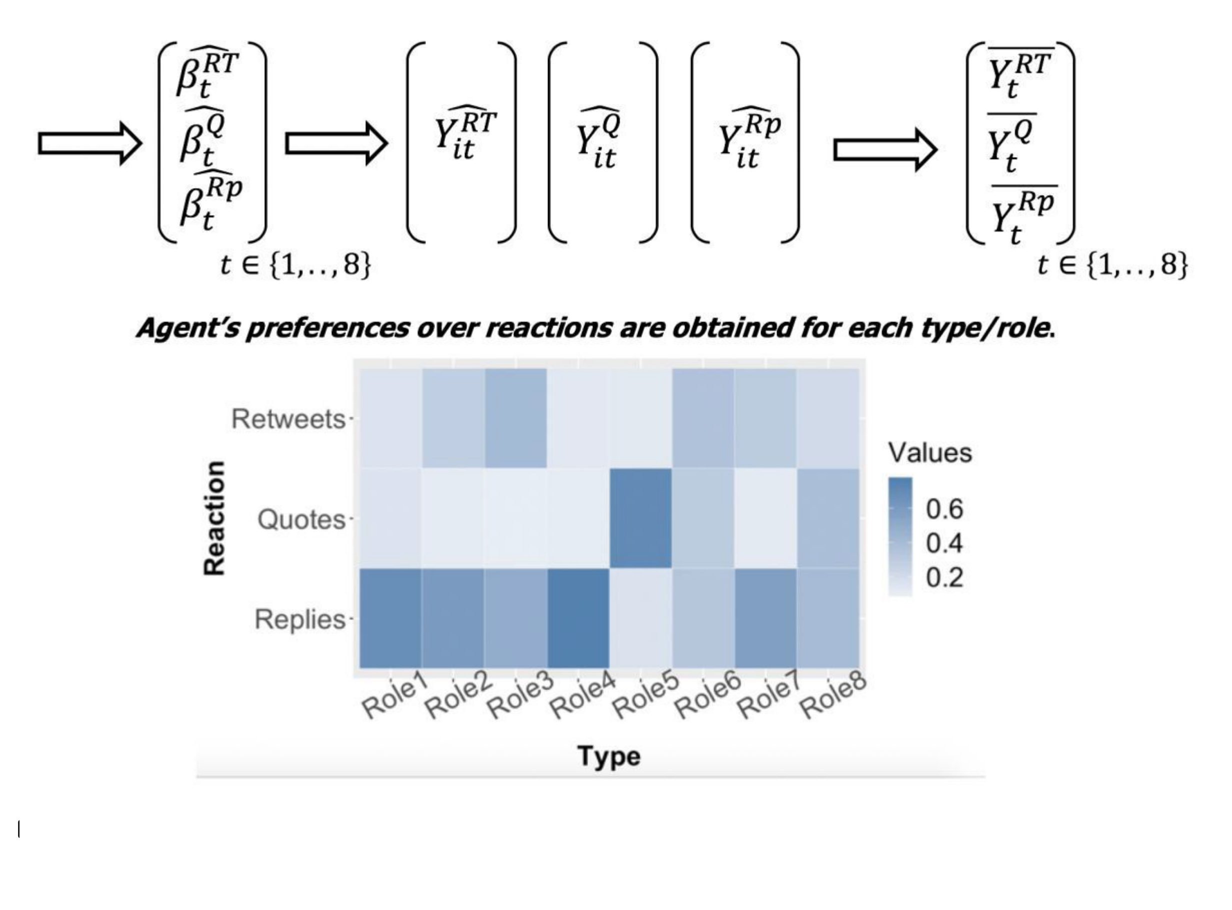


Figure 21. The Calculation of Preferences for Reactions in CP2.

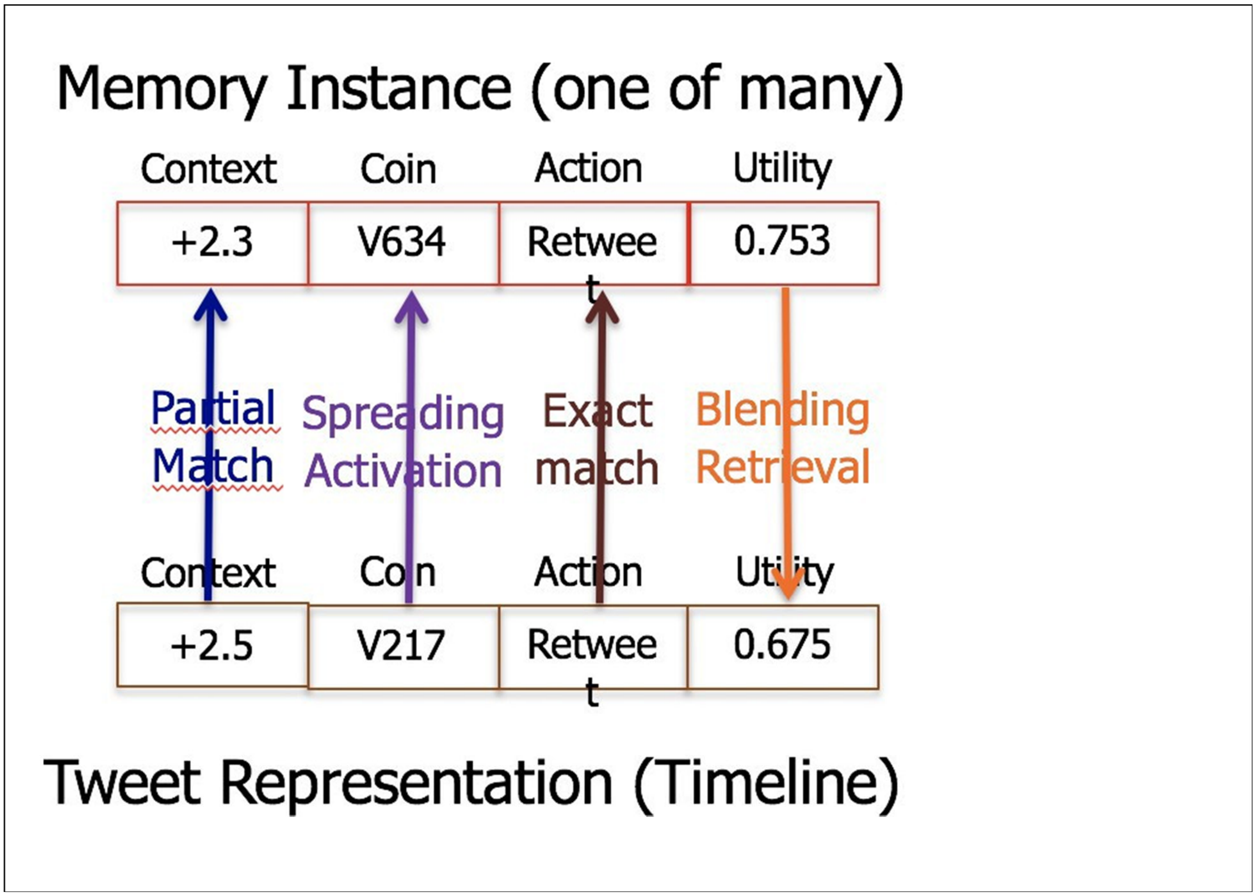


Figure 22. The Memory Instance in Act-R for CP2

Bayesian Top-Down Mode (GitHub only): The modeling approach for CP2 was similar in principle and motivation as that used in CP1 with the following important advances:

New model – BiDirectional Stacked Hawkes-Decay LSTMs

- Architecture followed from psychologically validated cognitive model of event knowledge
- Influence of hidden state for event prediction decays accorded to the kernel of a Hawkes process
- Each "stack" was one social medium -> event knowledge of the medium being treated as an extended mind.
- Multiple minds influenced each other via sharing of their temporally updating hidden layer

Exogenous information (e.g., cryptocurrency prices) could be included to influence event generation.

Figure 23 shows the main components of the modeling approach.

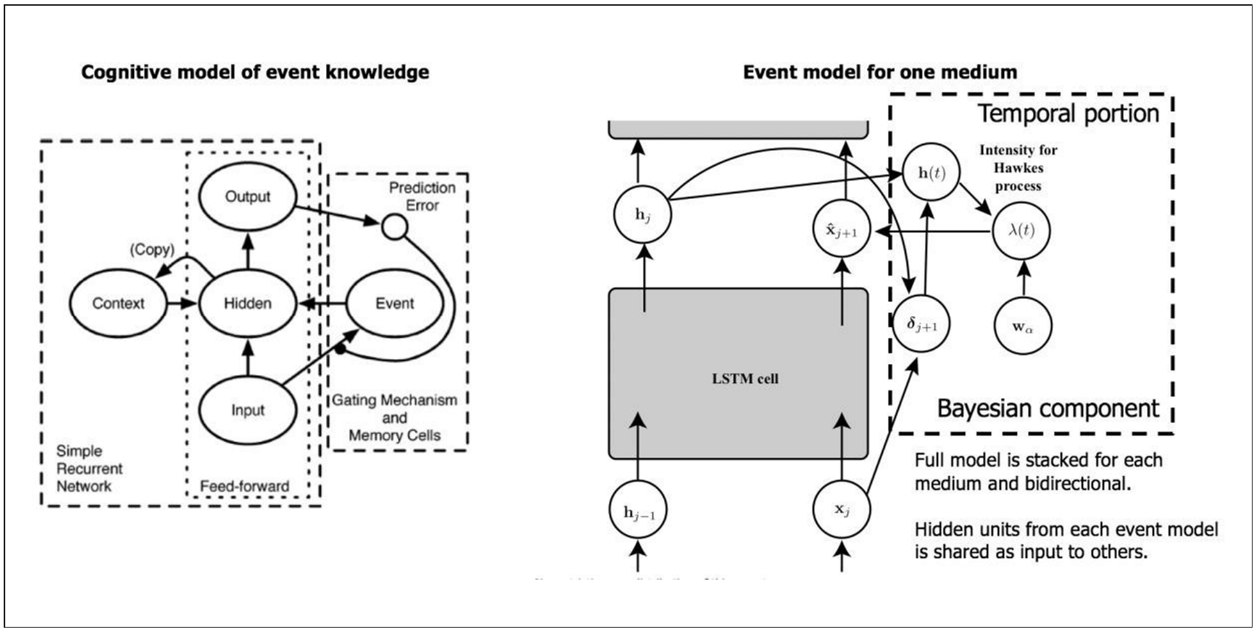


Figure 23. The Bayesian Top-Down Model Structure for CP2.

4.0 RESULTS AND DISCUSSION

For each challenge problem, the Testing and Evaluation (T&E) contractor provided a set of metrics that compared each of our simulations to ground truth. For each challenge, we present our results using these metrics. In particular, our analysis will focus on comparing our in-house baseline model to the cognitive and social theoretic models.

4.1 CP0

In this challenge, our analysis ingested the T&E metrics for each of our models and computed some comparison across the models. Our main goal was to compare our own baseline "Apoc" to the social cognition model and the cognitive architecture model. Figure 24 shows both bounded and unbounded metric scores for our models (by unbounded we mean metrics that had a max or min of inf or -inf). The primary comparison was between models labeled as "Soc_Cog" and "Cog_Arch" to "Mouse" (which was the baseline (also called Apoc)). Two points merit attention. First, the baseline model was generally similar across metrics to the cognitive architecture model (for both the bounded and unbounded metrics). Second, the social cognition model performed better than both baseline and cognitive architecture for the unbounded metrics and worse for the bounded metrics. The summary means (sd) for the models were:

Bounded: cog_arch = 0.74 (0.35); soc_cog = 0.54 (0.41); baseline = 0.67 (0.38)
UnBound: cog_arch = 0.61 (0.34); soc_cog = 0.64 (0.38); baseline = 0.51 (0.37)

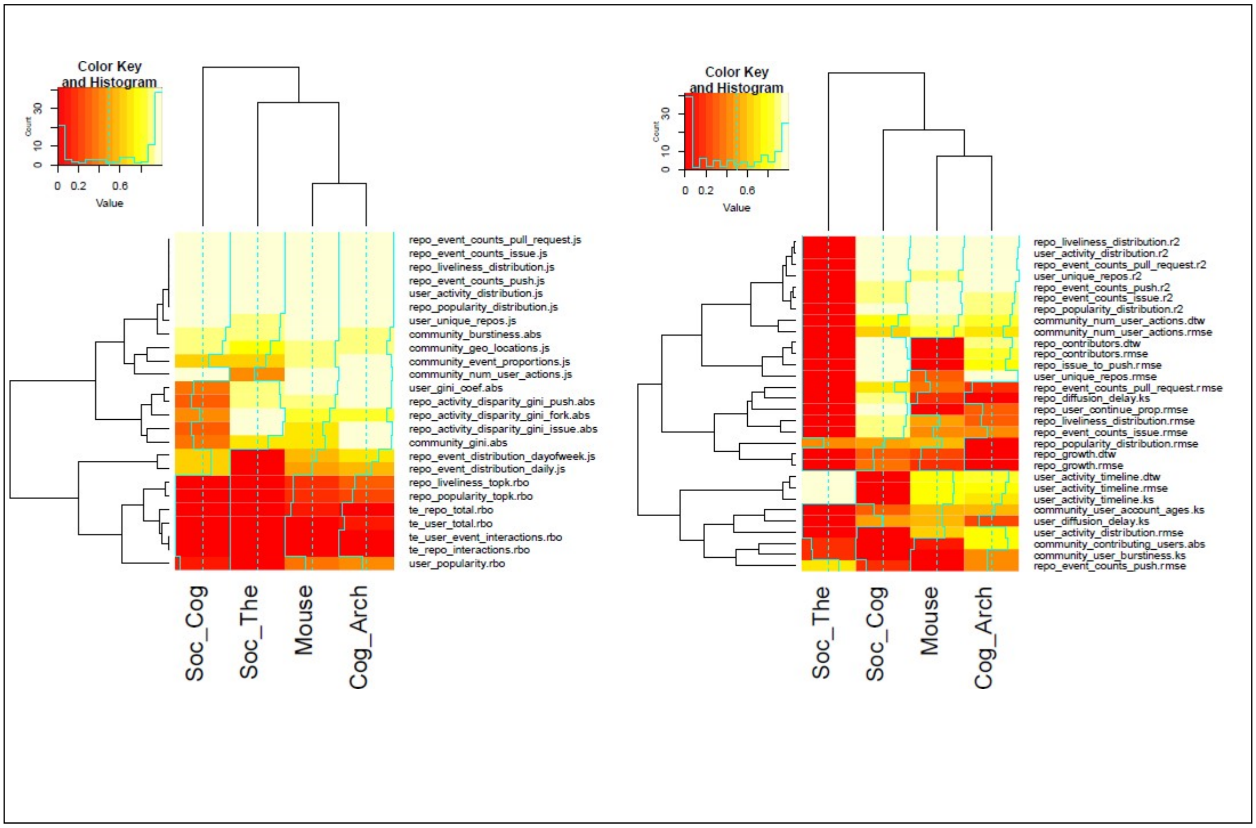


Figure 24. Visualization of the Comparison Between Baseline and Cognivive Models for CP0 (Bounded Metrics are Presented left; Higher Values Represent Less Error Between the Model and Ground Truth).

.

4.2 CP1

For this analysis, we developed an internal pipeline that allowed for running multiple replicates of our simulations and for computing the metrics developed by the T&E team. This allowed us to provide more rigor in our analysis. As for CP0, we focus on comparing the cognitive/social theoretic titration described in our methods to our baseline. Further, this pipeline afforded the capacity to improve our models from the CP0 performance, an important part of our model development.

Our first analysis was designed to compare the average performance of our cognitive/social theoretic models from CP0 to CP1. Figure 25 shows this comparison. In Figure 25, each point represents the average normalized metric performance for one simulation (error bars are + - two standard errors of the mean). The grey points represent the baseline statistical models; black points represent social/cognitive theory models. Two points merit mention. First, it is clear that there were general improvements from CP0 to CP1. Second, and more importantly, the social/cognitive outperformed the baseline in both CP0 and CP1.

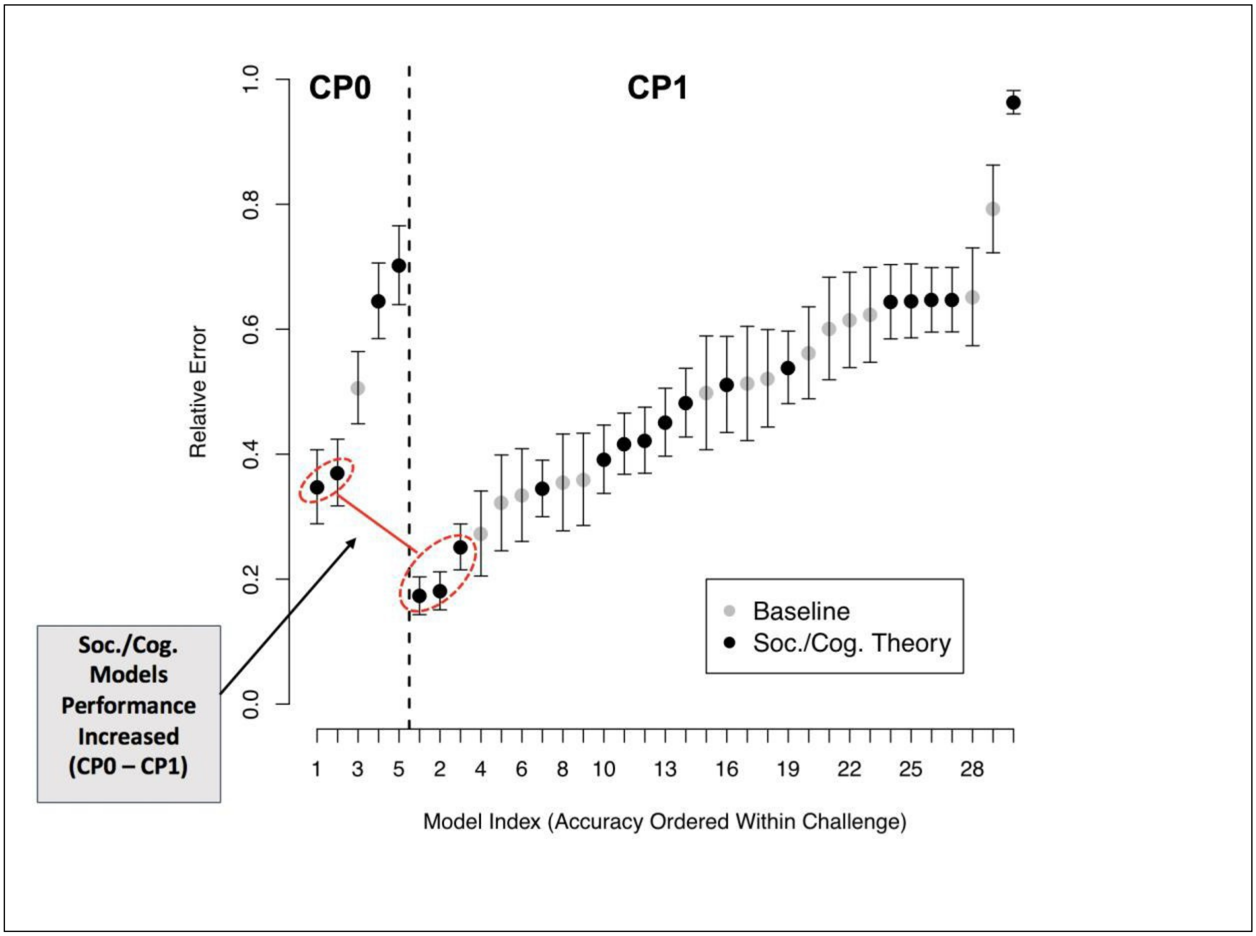


Figure 25. Comparison of Model Performance from CP0 to CP1.

The next analysis step was to determine if we could gain some improvements in our modeling approach within CP1. The performance period, after CP1, had some effort towards improvement of our modeling approach. For this, we used our internal simulation replicate/metric computation pipeline. Figure 26 shows this development. In Figure 26, each point represents the average normalized metric performance for one simulation (error bars are + - two standard errors of the mean; error bars removed for baseline model for readability). The grey points represent the baseline statistical models; black points represents social/cognitive theory models. Two points are relevant. First, after the CP1 challenge event, it is clear that we were able to improve the performance of the cognitive/social theoretic modeling effort. Second, the cognitive/social theoretic models performance was better than our internal baseline. A more detailed picture of the improvement period post CP1 is provided in Figure 27. This figure represents all replicates. The pink represent the baseline. The larger black points with error bars (one standard dev.) represent the average of the baseline model.

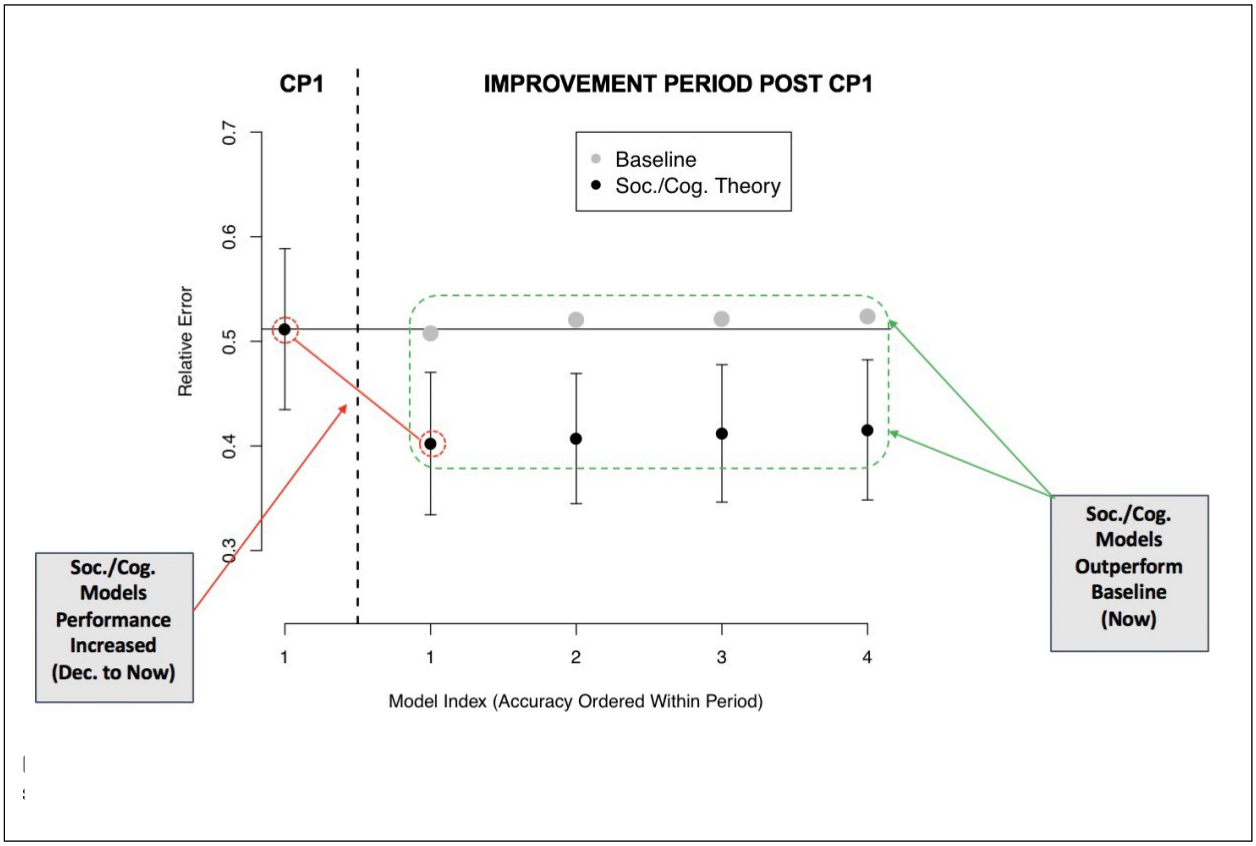


Figure 26. Comparison of Model Performance from CP1 Challenge Event to Further Developments in the Same Model.

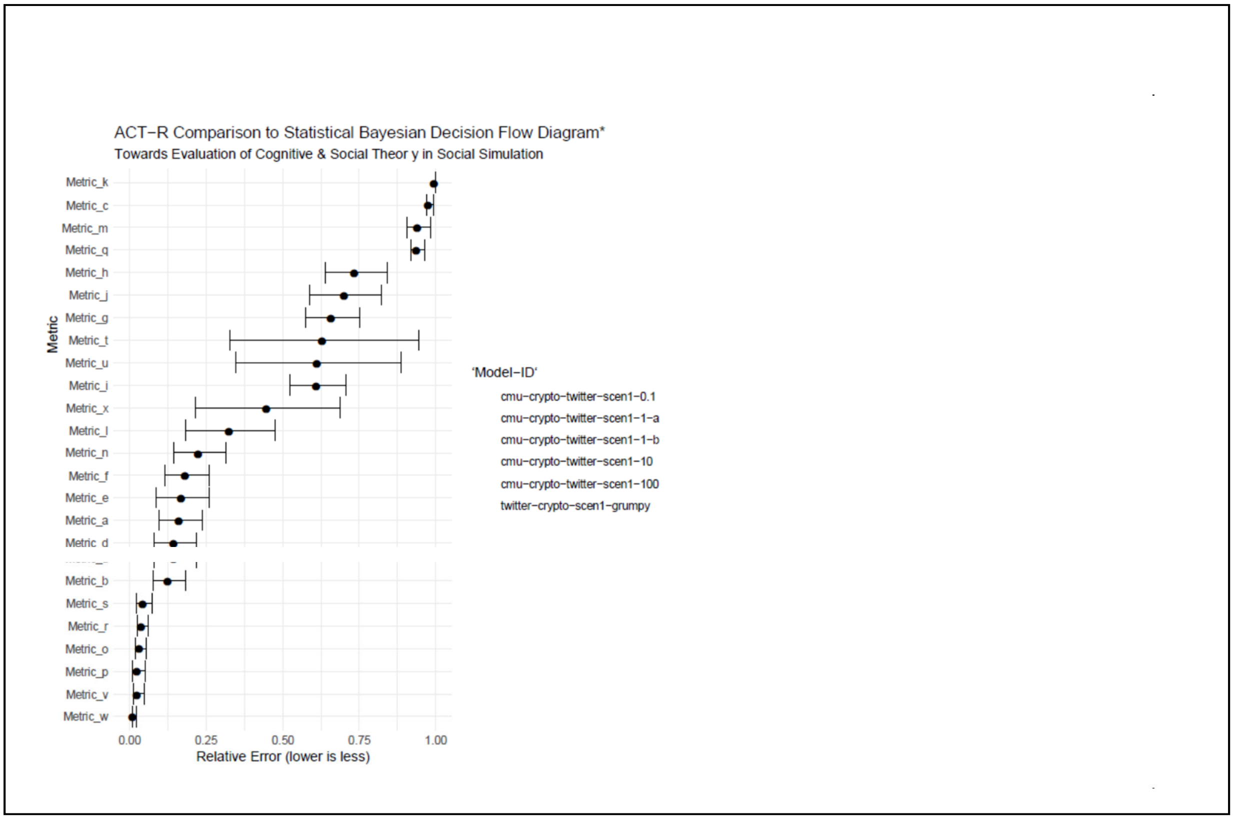


Figure 27. Comparison of Model Performance from CP1 Challenge Event to Further Developments in the Same Model with Details of Replicates.

4.3 CP2

In this challenge, our analysis ingested the T&E metrics for each of our models and computed an average comparison. We did not have the time, given the performance period and program goals to do an extensive analysis as we did for CP1. However, we did compare our own baseline BDFD model to the cognitive/social theoretic titration. Figure 28 shows this comparison. The first row shows the BDFD baseline model; the following rows show instantiations of the cognitive/social theory titration models. The baseline model had an average error of 0.36 while the mean for the cognitive/social theory modeling had a mean error of 0.28 (0.20 sd). The main finding, as in CP0 and CP1 is that our theoretic models outperformed the baseline models.

```
 [1,] "scenario2_grumpy-fast-1d_6ad07e50-aa26-4a8e-a0bd-78e6694ef3b7_metrics"       "0.36"
 [2,] "scenario2_moldy-grump-d-1_fb549c61-d521-4b1b-8e24-e50dfac5db06_metrics"      "0.27"
 [3,] "scenario2_moldy-grump-d-2_a18d7d60-ad91-4e86-9d5d-9890d5a3c4d0_metrics"      "0.24"
 [4,] "scenario2_moldy-grump-d-3_8af2666e-03ab-4b7f-945c-7017383ef888_metrics"      "0.3"
 [5,] "scenario2_moldy-grump-d-4_ad63199c-d782-4870-bc12-c85f3cd4075c_metrics"      "0.28"
 [6,] "scenario2_moldy-grump-n-1_9fbb826f-05a5-4ac5-b8c2-47c3f586a963_metrics"      "0.26"
 [7,] "scenario2_moldy-grump-n-2_71670f88-a334-4076-94f8-ca4da23828ec_metrics"      "0.28"
 [8,] "scenario2_moldy-grumpy-n-3_8a096919-908f-4bbf-b7ca-753aadc7eeae_metrics"     "0.27"
 [9,] "scenario2_moldy-grumpy-n-4_65d4f298-be02-436d-b859-e6be0d228764_metrics"     "0.3"
[10,] "scenario2_moldy-grumpy-stand-1_1c7d80b1-5f06-4d42-9101-1e980a9c4213_metrics" "0.3"
[11,] "scenario2_moldy-grumpy-stand-2_3b736e83-18ca-4d84-9185-14de391d1246_metrics" "0.25"
[12,] "scenario2_moldy-grumpy-stand-3_e83dcae7-d1fc-4583-a79b-c4acfdb293f8_metrics" "0.3"
[13,] "scenario2_moldy-grumpy-stand-4_36970f06-9d37-43eb-a4f1-c807413dd3f4_metrics" "0.27"
```

Figure 28. Comparison of Model Performance in Error in CP2 (yellow [first row] is Baseline BDFD)

5.0 CONCLUSIONS

The work conducted under the performance period was designed to build a technical platform and a set of methods that would to provide deep insight into the technical and theoretical best practices for and limitations of accurately scaling models of human cognition, perception, action and motivation to models of populations of cognitive agents that can accurately simulate online social behavioral phenomena. Central to our approach was the incorporation of substantive social behavioral theory into computational cognitive models of agents (to stand in as individual humans in our simulations). Our team developed and operated a technical platform, called *the Matrix*, and developed a conceptual prototype for a Bayesian top-down platform for these purposes.

The question of success is always difficult when endeavoring to cross disciplines especially social and behavioral sciences with computational sciences and large scale computing and software development. However, we think that the results presented above provide evidence, as proof-of-concept, that there is value in attempting to bridge the disciplines that are implicated in large human social systems. Some of the limitations are clear: 1) we have yet to dig really deeply into the relation between cognitive first principles, social theory and the kinds of generative causal links there are to larger-scale social structures and dynamics, 2) we did not link to neuroscience or genetics (but should, in a sense, dig all the way down).

6.0 REFERENCES

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2012). Human memory search as a random walk in a semantic network. In F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25 (pp. 3041-3049).

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. Psychological Review, 122(3), 558-569.

Anderson, S., Goeree, J., and Holt, C. (1998). A theoretical analysis of altruism and decision error in public goods games. Journal of Public Economics, 70, 297-323.

Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. Cognitive Science, 26, 85-112.

Austerweil, J. L., & Griffiths, T. L. (2013). A nonparametric Bayesian framework for constructing flexible feature representations. Psychological Review, 120, 817-851.

Austerweil, J. L., Brawner, S., Greenwald, A., Hilliard, E., Ho, M., Littman, M. L., MacGlashan, J., & Trimbach, C. (2016). The Impact of Outcome Preferences in a Collection of Non-Zero-Sum Grid Games. AAAI Spring Symposium 2016 on Challenges and Opportunities in Multiagent Learning for the Real World.

Austerweil, J. L., & Griffiths, T. L. (2011). Seeking confirmation is rational for deterministic hypotheses. Cognitive Science, 35, 499-526.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. Cognition, 113(3), 329-349.

Bergen, L., Levy, R. & Goodman N.D. (2016). Pragmatic Reasoning through Semantic Inference. Semantics and Pragmatics, 9, Article 20.

Bernheim B. D. (1984). Rationalizable strategic behavior. Econometrica 52(4), 1007–1028

Bhattacharyya, S., & Ohlsson, S. (2010). Social creativity as a function of agent cognition and network properties: A computer model. Social Networks, 32(4), 263-278.

Bornstein, G., Yaniv, I. (1998). Individual and Group Behavior in the Ultimatum Game: Are groups more 'rational' players? Experimental Economics, 1(1), 101-108.

Burkett, D., & Griffiths, T. L (2010). Iterated learning of multiple languages from multiple teachers. In Smith, A. M., Schouwstra, M., de Boer, B., and Smith, K. (Eds.) The Evolution of Language, 58-65.

Cabrera, S., Capra, C. M., Gómez, R. (2007). Behavior in one-shot traveler's dilemma games: model and experiments with advice, Spanish Economic Review 9 (2), 129-152

Camerer, C. (2003). Behavioral Game Theory. Experiments in Strategic Interaction, Princeton University Press, Princeton, New Jersey.

Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, *119*(3), 861-898.

Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties 1. American journal of Sociology, 113(3), 702-734.

Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. Cognition, 22(1), 1–39.

Ehret, P. J., Monroe, B. M., & Read, S. J. (2015). Modeling the Dynamics of Evaluation: A Multilevel Neural Network Implementation of the Iterative Reprocessing Model. Personality and Social Psychology Review, 19(2), 148-176.

Erev, I., Ert, E., Roth, A. E., Haruvy, E., Herzog, S., Hau, R., Hertwig, R., Stewart, T., West, R., Lebiere, C. (2010). A choice prediction competition, for choices from experience and from description. Journal of Behavioral Decision Making 23(1), 15-47.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. Science, 336, 998.

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. Cognitive Psychology, 75, 80-96.

Freeman, J. B., & Ambady, N. (2011). A dynamic interactive theory of person construal. Psychological Review, 118, 247-281

Gloeckner, A., Hilbig, B. E., & Jekel, M. (2014). What is adaptive about adaptive decision making? A parallel constraint satisfaction account. Cognition, 133, 641-666.

Gonzalez, C., Ben-Asher, N., Martin, J. M., & Dutt, V. (2015). A cognitive model of dynamic cooperation with varied interdependency information. Cognitive science, 39(3), 457-495.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. Trends in Cognitive Sciences, 20(11), 818-829.

Graf, C., Degen, J., Hawkins, R. X. D. & Goodman N.D. Animal, dog, or Dalmatian? Level of abstraction in nominal referring expressions. (2016). In Proceedings of the Thirty-Eighth Annual Conference of the Cognitive Science Society.

Hawkins, R. X. D., Frank, M. C. & Goodman N. D. (2017). Convention-formation in iterated reference games. In Proceedings of the Thirty-Ninth Annual Conference of the Cognitive Science Society.

Ho, M. K., Littman, M. L., MacGlashan, J., Cushman, F. & Austerweil, J. L. (2016). Showing versus doing: Teaching by demonstration. In Advances in Neural Information Processing System, 29, 3027-3035.

Ho, M. K., MacGlashan, J., Hilliard, E., Trimbach, C., Brawner, S., Gopalan, N., Greenwald, A., Littman, M. L., & Austerweil, J. L. (2016). Feature-based Joint Planning and Norm Learning in Collaborative Games. In Proceedings of the 38th Annual Meeting of the Cognitive Science Society.

Jern, A., Chang, K.-M. K., Kemp, C. (2014). Belief polarization is not always irrational. Psychological Review, 121(2), 206-224.

Juvina, I., Lebiere, C., & Gonzalez, C. (2015). Modeling trust dynamics in strategic interaction. Journal of applied research in memory and cognition. 4(3): 197-211.

Juvina, I., Saleem M., Gonzalez, C., & Lebiere, C., (2012). Reciprocal Trust Mediates Deep Transfer of Learning Between Games of Strategic Interaction. Organizational Behavior and Human Decision Processes. Special Issue in Social Dilemmas. http://dx.doi.org/10.1016/j.obhdp.2012.09.004

Kao, J.T., & Goodman, N.D. Let's talk (ironically) about the weather: Modeling verbal irony. (2015). In Proceedings of the Thirty-Seventh Annual Conference of the Cognitive Science Society.

Kao, J.T., Wu, J., Bergen, L. & Goodman, N.D. Nonliteral understanding of number words. (2014). Proceedings of the National Academy of Sciences, 111(33), 12002-12007.

Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. Current opinion in neurobiology, 13(2), 150-158.

Klayman, J. (1995). Varities of confirmation bias. In J. Busemeyer, R. Hastie, & D. Medin (Eds.), Psychology of learning and motivation: Vol. 32. Decision making from a cognitive perspective, pp. 365-418. New York: Academic Press.

Kleiman-Weiner, A., Ho, M. K., Austerweil, J. L., Littman, M. L., & Tenenbaum, J. B. (2016). Coordinate to cooperate or compete: Abstract goals and joint intentions in social interaction. In Proceedings of the 38th Annual Meeting of the Cognitive Science Society.

Korkmaz, G., C.J. Kuhlman, S.S. Ravi, and F. Vega-Redondo. 2016. Approximate Contagion Model of Common Knowledge on Facebook. In Proceedings of the 27th ACM Conference on Hypertext and Social Media (HT' 16). ACM, New York, NY, USA, 231-236.

Kunda, Z., & Thagard, P. (1996). Forming impressions from stereo- types, traits, and behaviors: A parallel-constraint-satisfaction theory. Psychological Review, 103, 284-308.

Lassiter, D., & Goodman, N. D. (2017). Adjectival vagueness in a Bayesian model of interpretation. *Synthese*, *194*(10), 3801-3836.

Lazer, et al. (2009). Computational social science. Science, Vol. 323, Issue 5915, pp. 721-723

Lebiere, C., Stocco, A., Reitter, D., & Juvina, I. (2010). Scaling Up High-Fidelity Cognitive Modeling to Real-World Applications. In Proceedings of NATO Workshop on Human Modeling for Military Application. Amsterdam, NL: October 18-20, 2010.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature, 521, 436-444.

Lejarraga, T., Lejarraga, J., & Gonzalez, C. (2014). Decisions from experience: How groups and individuals adapt to change. Memory & cognition, 42(8), 1384-1397.

Lewandowsky, S., & Oberauer, K. (2016). Motivated rejection of science. Current Directions in Psychological Science, 25(4), 217-222.

Ma, W. J. (2012). Organizing probabilistic models of perception. Trends in Cognitive Sciences, 16(10), 511-518.

Marsella, S. C., Pynadath, D. V., & Read, S. J. (2004). PsychSim: Agent-based modeling of social interactions and influence. Paper in the Proceedings of the 6th International Conference on Cognitive Modeling, Carnegie Mellon University, Pittsburgh, PA.

Minsky, M. and Papert, S. (1969). Perceptrons. Cambridge, MA: MIT Press.

Monroe, B. M., & Read, S. J. (2008). A General Connectionist Model of Attitudes and Attitude Change: The ACS (Attitudes as Constraint Satisfaction) Model. Psychological Review, 115, 733–759.

Monroe, W., Hawkins, R. X. D., Goodman, N. D. &Potts, C. (2017). Colors in Context: A Pragmatic Neural Model for Grounded Language Understanding. Transactions of the Association for Computational Linguistics, vol. 5, pp. 325–338.

Navarro, D. J., & Perfors, A. (2011). Hypothesis generation, the positive test strategy, and sparse categories. Psychological Review, 118, 120-134.

Orr, M. G., Ziemer, K. & Chen, D. (2017). Systems of behavior and population health. In (Eds El-Sayed & Galea). Systems Science and Population Health. Oxford Univ. Press, New York.

Orr, M. G., & Plaut, D. C. (2014). Complex systems and health behavior change: insights from cognitive science. American journal of health behavior, 38(3), 404-413.

Orr, Mark G., Roxanne Thrush, and David C. Plaut. "The theory of reasoned action as parallel constraint satisfaction: towards a dynamic computational model of health behavior." PloS one 8.5 (2013): e62490.

Orr, M. G., Chen, D. (2018). Computational models of health behavior. In Vallacher, R., Nowak, A., & Read, S. (Eds.), Computational Models in Social Psychology. Psychology Press/Routledge: New York.

Quattrociocchi, W. (2017). Inside the echo chamber. Scientific American. 316(4), 60-63.

Read, S. J., & Miller, L. C. (1993). Rapist or "regular guy": Explanatory coherence in the construction of mental models of others. Personality and Social Psychology Bulletin, 19, 526-540.

Read, S. J., & Miller, L. C. (1994). Dissonance and Balance in belief systems: The promise of parallel constraint satisfaction processes and connectionist modeling approaches. In R. C. Schank & E. J. Langer (Eds.), Beliefs, reasoning, and decision making: Psycho-logic in honor of Bob Abelson (pp. 209-235). Hillsdale, NJ: Erlbaum.

Read, S. J., & Monroe, B. M. (2019). Modeling cognitive dissonance as a parallel constraint satisfaction network with learning. In E. Harmon-Jones (Ed.), Cognitive dissonance: Reexamining a pivotal theory in psychology, pgs. 197–226.

Reitter, D., & Lebiere, C. (2012). Social cognition: Memory decay and adaptive information filtering for robust information maintenance. In Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI-12).

Reitter, D. & Lebiere, C. (2011). How groups develop a specialized domain vocabulary: A cognitive multi-agent model. Journal of Cognitive Systems Research, 12(2), 175-185.

Reitter, D., & Lebiere, C. (2011). Towards cognitive models of communication and group intelligence. In Proceedings of the 33rd Annual Meeting of the Cognitive Science Society

Reitter, D., & Lebiere, C. (2010). On the influence of network structure on language evolution. In R. Sun, editor, Proceedings of Workshop on Cognitive Social Sciences: Grounding the Social Sciences in the Cognitive Sciences (at Cognitive Science: CogSci 2010), Portland, Oregon.

Reitter, D., & Lebiere, C. (2010). Accountable Modeling in ACT-UP, a Scalable, Rapid-Prototyping ACT-R Implementation. In Proceedings of the 2010 International Conference on Cognitive Modeling.

Romero, D. M., Meeder, B., & Kleinberg, J. (2011, March). Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In Proceedings of the 20th international conference on World wide web.

Romero, O., & Lebiere, C. (2014). Simulating Network Behavioral Dynamics by using a Multi-agent approach driven by ACT-R Cognitive Architecture. In Proceedings of the Behavior Representation in Modeling and Simulation Conference (BRIMS-2014).

Rosenblatt, F. (1962). Principles of neurodynamics. New York: Spartan.

Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning representations by back-propagating errors. Nature, 323, 533-536.

Sanborn, A. N., Griffiths, T. L., & Navarro D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. Psychological Review, 117(4), 1144-1167.

Shoda, Y., LeeTiernan, S., & Mischel, W. (2002). Personality as a dynamical system: Emergence of stability and distinctiveness from intra and interpersonal interactions. Personality and Social Psychology Review, 6(4), 316-325.

Stahl, Dale and Wilson, Paul. (1995). On Players' Models of Other Players: Theory and Experimental Evidence. Games and Economic Behavior. 10(1), 218-254.

Taatgen, N. A. & van Rijn, H. (2008). ACT-R Tutorial. Proceedings of the 30th Annual Meeting of the Cognitive Science Society.

Tang, Y., Lebiere, C., Sycara, K., Morrison, D., Lewis, M., & Smart, P. (2015). Information Sharing for Collective Sensemaking. In Proceedings of the Hawaii International Conference on System Sciences (HICSS-49).

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to Grow a Mind: Statistics, Structure, and Abstraction. Science, 331, 1279-1285.

Thagard, P. (1989). Explanatory coherence. The Behavioral and Brain Sciences, 12, 435-467.

Tran, D., Kucukelbir, A., Dieng, A. B., Rudolph, M., Liang, D., & Blei, D. M. (2016). Edward: A library for probabilistic modeling, inference, and criticism. arXiv preprint arXiv:1610.09787.

Van Overwalle, F., & Heylighen, F. (2006). Talking nets: A multiagent connectionist approach to communication and trust between individuals. Psychological Review, 113(3), 606.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and Done? Optimal Decisions From Very Few Samples. Cognitive Science, 38(4), 599-637.

Watts, D. J. (2002). A simple model of global cascades on random networks. Proceedings of the National Academy of Sciences, 99(9), 5766-5771.

West, R. L., Lebiere, C. & Bothell, D. J. (2006). Cognitive architectures, game playing and human evolution. In Sun, R. (Ed) Cognition and Multi-Agent Interaction: From Cognitive Modeling to Social Simulation. NY, NY: Cambridge University Press. Pp. 103-121.

Whalen, A., & Griffiths, T. L. (2017). Adding population structure to models of language evolution by iterated learning. Journal of Mathematical Psychology, 76, 1-6.

Zhang, Y., & Leezer, J. (2010). Simulated human-like decisions in a memory-based agent model. Computational and Mathematical Organization Theory, 16, 373-399.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMNS

AAAI	Association for the Advancement of Artificial Intelligence
AAMAS	Autonomous Agents and Multi-Agent Systems
ACM	Association for Computing Machinery
ACS	American Chemical Society
ACT-R	Adaptive Control of Thought—Rational
ACT-UP	A toolkit implementation of ACT-R
AFRL	Air Force Research Lab
aPFC	Anterior prefrontal cortex
API	Application Programming Interface
BDFD	Bayesian Decision Flow Diagram
BRIMS	Behavior Representation in Modeling and Simulation
C	a general-purpose, procedural computer programming language
CM	Cognitive Model
CMU	Carnegie Mellon University
COTR	Contracting Officer's Technical Representative
CP0	Challenge Problem 0
CP1	Challenge Problem 1
CP2	Challenge Problem 2
CPU	Central Processing Unit
CVE	Common Vulnerabilities and Exposures
DAG	Directed Acyclic Graph
DARPA	Defense Advanced Research Projects Agency
DLPFC	Dorsolateral Prefrontal Cortex
DOOCN	Dynamics of On and Off Complex Networks
HICSS	Hawaii International Conference on System Sciences
IBL	instance-based learning
IBLT	instance-based learning theory
IEEE	Institute of Electrical and Electronics Engineers
IHMC	Institue for Human Machine Cognition
IPC	interprocess communication
JSON	JavaScript Object Notation
LDE	lower dimensional embedding
LMDB	Lightning Memory-Mapped Database
LSTM	Long Short Term Memory
MA	Massachusetts
NATO	North American Treaty Organization
NY	New York
PA	Pennsylvania
R	a programming language and free software environment for statistical computing and graphics
RSA	Rational Speech Act

SBP-BRIMS	Intl. Conf. on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation
SDU	Situation, decision, utility memory instance
SDU/U	Situation, decision, utility memory instance with observed utility
SQL	Structured Query Language
STM	Social Theory Model
T&E	Test & Evaluation
TA1	Technical Area 1
TCP/IP	transmission control protocol/Internet protocol
USC	University of Southern California
UVA	University of Virginia
VLPFC	ventrolateral prefrontal cortex
VT	Virginia Tech

APPENDIX A – Publications and Presentations

Books, Articles, and Proceedings

- Bhattacharya, P., Ekanayake, S., Kuhlman, C.J., Lebiere, C., Morrison, D., Swarup, S., Wilson, M.L., and Orr, M.G.; "The Matrix: An Agent-Based Modeling Framework for Data Intensive Simulations"; *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019),* pp. 1635-1643; May 2019
- Swarup, S.; "Adequacy: What Makes a Simulation Good Enough?"; Paper presented at the Spring Simulation Conference; May 2019
- Orr, M.G.; "Multi-Scale Resolution of Human Social Systems: A Synergistic Paradigm for Simulating Minds and Society"; *Social-Behavioral Modeling for Complex Systems*; edited by Davis, P.K., O'Mahony, A., and Pfautz, J.; John Wiley & Sons, 2019, pp. 697-710.
- Swarup, S., Marathe, A., Marathe, M.V., and Barrett, C.L.; "Simulation Analytics for Social and Behavioral Modeling"; *Social-Behavioral Modeling for Complex Systems*; edited by Davis, P.K., O'Mahony, A., and Pfautz, J.; John Wiley & Sons, 2019, pp. 617-632.
- Orr, M.G., Lebiere, C., Stocco, A., Pirolli, P., Pires, B., and Kennedy, W.G.; "Multi-scale resolution of neural, cognitive and social systems"; *Computational and Mathematical Organization Theory,* 25 (2019): 4-23.
- Orr, M.G., Lebiere, C., Stocco, A., Pirolli, P., Pires, B., and Kennedy, W.G.; "Multi-scale Resolution of Cognitive Architectures: A Paradigm for Simulating Minds and Society"; *Social, Cultural, and Behavioral Modeling;* edited by Thomson, R., Dancy, C., Hyder, A., and Bisgin, H.; Springer, 2018, pp. 3-15. [Presented at the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS); July 2018]

Conference Presentations

- Bhattacharya, P., Ekanayake, S., Kuhlman, C.J., Lebiere, C., Morrison, D., Swarup, S., Wilson, M.L., and Orr, M.G.; "The Matrix: An Agent-Based Modeling Framework for Data Intensive Simulations"; Talk presented at the International Conference on Autonomous Agents and Multiagent Systems (AAMAS); May 2019
- Orr, M.G., Lebiere, C., Stocco, A., Pirolli, P., Pires, B., and Kennedy, W.G.; "Multi-Scale Resolution of Cognitive Architectures: Connecting Levels between the Newell Bands"; Paper presented at the Association for the Advancement of Artificial Intelligence (AAAI) Fall Symposium on A Common Model of Cognition; October 2018
- Bhattacharya, P.; "The Matrix: An Agent-Based Modeling Framework for Complex Systems and Data Intensive Simulations"; Invited talk at the Dynamics On and Of Complex Networks Workshop (DOOCN-XI) at the Complex Systems Conference (Thessaloniki, Greece); September 2018
- Orr, M.G.; Panel presentation on Miscellaneous Topics; Invited panel presentation for "Physics of Information Processing" at the 3rd International Workshop on Social Sensing (Orlando, Florida); April 2018

Awards

- Best Paper for the International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation (SBP-BRiMS) 2018

Press and Communications

- Lebiere, C.; "CMU researchers are building a model to predict human behavior. It could save lives one day," Interviewed by Linder, C.; Pittsburgh Post-Gazette; February 6, 2018 [https://www.post-gazette.com/business/tech-news/2018/02/06/cmu-cognitive-architecture-social-sim-act-r-virginia-tech-darpa-pittsburgh/stories/201801180005]
- Orr, M., Pfautz, J., Moody, J., and Barrett, C.; "Virginia Tech team leads federal effort to forecast the flow of information online"; Interviewed by Rosplock, D.; Virginia Tech Daily; December 7, 2017 [https://vtnews.vt.edu/articles/2017/11/bi-forecasting-info-flow-online0.html]