Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

IMPACTS OF BEHAVIORAL MODELING ASSUMPTIONS FOR COMPLEX ADAPTIVE SYSTEMS: AN EVALUATION OF AN ONLINE DATING MODEL

Andrew J. Turner

Vienna, VA 22180, USA

Jennifer Kuczynski

MITRE
7525 Colshire Dr
McLean, VA 22102, USA

ABSTRACT

This paper investigates the impact of behavioral modeling assumptions for a Complex Adaptive System (CAS) model. We hypothesize that behavioral models can be overconfident in their predictions due to the challenges of modeling behavior. Supporting this hypothesis, this paper discusses the challenges of modeling behavior, presents a CAS example problem to design an online dating app, models the dating app as a CAS, and investigates the impacts of different behavioral models on the design. This paper shows how similar behavioral models can have a significant impact on the simulation results. This paper highlights the challenge our community faces moving forward with valid behavioral models. Finally, we call for the community at large to address these challenges by collaboratively researching and comparing behavioral models so as to guide future modelers.

1 INTRODUCTION

Behavioral modeling presents a major challenge to the modeler, especially in regards to human behavior. In part, this is due to the difficulty in understanding and modeling human behavior, limited time and resources to explore various behavior models, and limited options for validating these behavioral models. Many models face this challenge, but the field of Complex Adaptive Systems (CAS) should take lead on addressing these challenges due to the community's experience with and core dependency on behavioral models.

The variability of human behavior and our difficulty in understanding it, creates the greatest challenges to behavioral modeling. Scientists have long attempted to document and understand human behavior, but it has remained an elusive task. Psychology may be the oldest of the related sciences, but has faced recent inability to replicate classic psychological studies (Pashler and Wagenmakers 2012). In the military domain, the modeling of human behavior for anti-terrorism and stability operations has increased due to the wars in Afghanistan and Iraq (Lucas et al. 2003; Raczynski 2004; Hausken and Zhuang 2011), however there is limited evidence of any effective application or resolution of these complex challenges as a result of these studies. Our understanding of and modeling of human behavior is still in its infancy, and the failure to understand the system to be modeled leads to limited validation and potentially false conclusions.

Like any other project, modeling faces resource and time constraints. Building a valid model requires the modeler to identify how much of the real-world CAS must be modeled (i.e., breadth) and how much detail must be modeled (i.e., depth or fidelity). Commonly employed methods for behavioral representation are if-then conditionals, decision trees, probability models, and neural networks. Currently, it is unclear which methods are best for representing which behaviors, for little study has occurred on comparative behavioral model representation. Due to the limits on resources and time, modelers often employ only one behavioral representation. Both the singular behavioral representation and uncertainty of the proper model representation leads to more emphasis being placed on operational validation between the computerized model and the problem entity (Sargent 2013). This requires real world data from the problem entity.

Validation of models require a comparison to the real world to determine if the model is sufficiently accurate to answer the questions asked of it. The options for validating behavioral models are limited due to several factors. First, there is limited real world data that can be gathered on the behavior of complex systems (i.e., the systems are unobservable). Second, the data that is gathered is often subjective rather than objective, leading to greater potential for false validation (Turner et al. 2012). Finally, the modeled systems may not be static across time, location, or culture. For example, the data gathered for validation 20 years ago may not be representative of the system today. Similarly, the data gathered in western societies may not be representative of the rest of the population (Henrich et al. 2010). Therefore, modelers face significant challenges validating their behavioral models.

We claim that inadequately addressing these challenges results in overconfidence in the accuracy and validity of behavior models. This paper aims to illuminate the sensitivity of behavior modeling to simulation output measures. This is accomplished through the modeling of an online dating app where we ask how the dating app should be structured for user satisfaction. This example problem highlights the issues the community faces moving forward with valid behavioral models. We do not claim to have answers to the challenges of validation, but instead call for the CAS community to work towards a better understanding of CAS modeling, when to use specific behavioral representation models, and validation approaches for CAS models. We hope in the future there will be more comparative analysis of different behavioral models and guidance to new modelers on accepted state of the art of modeling particular behaviors than we currently observe.

The remainder of this paper is as follows. Section 2 introduces the example problem. In Section 3, the background of online dating services is discussed. Using the example problem and the background, Section 4 presents related models from the literature and presents our conceptual model. Section 5 discusses the validation of the models. Section 6 presents the analysis of the various behavior models. Finally, the paper concludes with a discussion of our findings and a call to the CAS community to address the challenges of modeling.

2 EXAMPLE PROBLEM

Our example problem is to design a dating app that provides its users with an overall satisfying experience. We define user satisfaction as the user's success at finding a partner. To capture this, we measure the number of committed couples formed and the number of users left without finding a committed partner. We hypothesize that a system that presents too many options to a user will lead to a lower overall user satisfaction. According to Schwartz (2004), "As the number of options continues to proliferate, making an exhaustive investigation of the possibilities impossible, concern that there may be a better option out there may induce you to anticipate the regret you will feel later on, when that option is discovered, and thus prevent you from making a decision at all." Some people have a fear of missing out on perfection, and are reluctant to give up on their choices. By clinging to the availability of options, either no choice is ever made or their anxiety and stress lead to dissatisfaction in their choices.

The requirement of our example problem is to provide a simplified form by which its application is representative of the general class of problem. For this, we need the problem to be defined as a CAS. While there is no one commonly accepted definition of CAS, there are several popular definitions offered in the literature. According to Mitchell (2009), "A complex system is one in which large networks of components with no central control and simple rules of operation give rise to complex collective behavior, sophisticated information processing, and adaptation via learning or evolution. In complex adaptive systems, adaptation plays a large role." Macal and North (2010) state that complex systems are composed of interacting, autonomous components, and CAS have the added benefit for agents to adapt (at either the individual or population levels). Szabo et al. (2014) state that "Complex systems often exhibit behavior that cannot be reduced only to the behavior of their individual components, and component interactions often result in new and unexpected properties." Each definition has themes in common: independent, components that interact with each other, adapt, and are not centrally controlled. We classify the online dating system as a

CAS because it exhibits each of these characteristics. In the system, each user is independent, interacts with other users in their own self-interest, updates their strategies, and is free from centralized control.

3 BACKGROUND ON DATING APPS

Online dating has become popular in recent years. One-in-ten Americans are using or have used an online dating platform, and approximately half of Americans know someone who has used online dating (Smith and Anderson 2014). As with the rising participation in online dating, so has risen a plethora of online dating platforms. There exists a wide variety in platform designs that claim to be best in matching people with compatible partners. Before designing another platform, it's important to know the variety that exists in the market. Differences exist at every stage of a user's interaction with an online dating platform. These differences include: how the user enters the platform, what information is required on their profile, whose profiles they are allowed to view, how many profiles they can view at one time, how messaging between users is handled, and how users are removed from the platform.

Online dating platforms take two major approaches for accepting people into their system. The first is an open approach in which the user is allowed to join the service at any time (e.g., OkCupid, Hinge, Coffee Meets Bagel, Tinder). The other approach is to limit access to the platform through a privileged access mechanism. This privileged access can be granted via an invite system (e.g., The League, The Inner Circle) or an application process (e.g., Raya, Luxy).

The information required of a user to create an account can range from a minimal description to extensive essays and quizzes. On the minimal side, apps like Tinder require a picture, age, and a few sentences. On the extensive side, apps like OkCupid require multiple fields for written response (e.g., self-summary, what you're doing with your life), multiple attributes (e.g., Height, Job, Income), and a quiz to determine match potential.

Designs differ greatly on how users see potential matches (i.e., the pool they are allowed to view at any one time). On some apps, one gets to scroll through all available profiles (e.g., OkCupid). Other apps show the user a number of algorithm-selected profiles per unit of time (e.g., Coffee Meets Bagel). Finally, others show a different number of profiles based on the user's app subscription. Here, differing subscription levels provide the user more (or better) matches (e.g., The League, Tinder).

How users message one another has the greatest variety across apps. Some allow all users to send a message to all other users, with no limits on the number of messages sent (e.g., OkCupid). Most other apps allow messages only after two users match; a process where the two select each other. Only after users are matched is messaging between parties allowed (e.g., Tinder, Coffee Meets Bagel, Hinge, Tinder). Constraints may exist on matches or messages allowed per time period. Other apps place additional restrictions such as a "women message first" rule (e.g., Bumble).

Ultimately, users will eventually leave the apps. They may leave the app of their own volition, due to having found a long-term partner or having become frustrated and wanting to take a break. They may also be forced to leave due to inappropriate behavior, lack of activity, or an expiration of their subscription.

As the market becomes oversaturated with online dating app options and the multitude of ways in which matching can be achieved, one wonders the optimal approach to matching. Our model will investigate one of the variables, which is the size of the dating pool in which the user is presented.

4 MODEL DEVELOPMENT

4.1 Literature Review of Models

Common modeling paradigms used for CAS include complex networks, system dynamics, and agent-based modeling. Complex network modeling uses nodes and edges in a mathematical graph to represent the connectivity and behavior of a system. It is applied to CAS problems across various scientific domains including computer science (Barabási and Albert 1999), chemistry (Guimera et al. 2004), and behavior science (Centola 2010). System dynamic modeling represents systems using stocks, flows, and feedback loops where the events and entities are aggregated (Borshchev and Filippov 2004). System dynamic models have been applied to a variety of disciplines including business processes (Sterman 1992; Borshchev and Filippov 2004), healthcare (Brailsford et al. 2004), and safety (Bouloiz et al. 2013; Cooke 2003). Agent based modeling is a bottom-up approach to modeling CASs, focusing on modeling autonomous agents, their behaviors, and their interactions with their environment (Macal and North 2007). Agent Based Models (ABM) are traditionally used for decentralized systems with emergent behavior, such as, biological systems (flocking behavior (Reynolds 1995), immune system response (Folcik and Orosz 2006)), social systems (e.g., urban crime analysis (Groff et al. 2018), social segregation (Schelling 1978)), and traffic/crowd simulations (Pan et al. 2007).

Before selecting a modeling paradigm, we reviewed the literature to find a precedent for modeling online dating apps. No papers were found that model an online dating app; however, a number were found that model the mate choice problem for humans and other animals. Cited as the earliest application of ABM to the mate choice problem is Kalick and Hamilton (1986), who apply their model to address the matching paradox (i.e., partners display correlation in attractiveness yet prefer highest attractive partner). They were able to show that agents with preference for most attractive mates can lead to correlated attractiveness as an emergent property. A few key design choices of their model include maintaining a constant number of agents (i.e., replacing the agents as the pair off), a single attractiveness attribute, random pairings for dates, and three probability functions for accepting a mate. These are shown in (1) and (2), with the third function being their average.

$$P = \left(\frac{A_j}{k}\right)^{C\lambda} \tag{1}$$

$$P = \left(\frac{k-|A_i-A_j|}{k}\right)^{C\lambda} \tag{2}$$

Such that:
k: maximum attractiveness (Kalick and Hamilton assumed 10)
A_i: Attractiveness perception of self
A_j: Attractiveness of potential match
C: Baseline choosiness (Kalick and Hamilton assumed 3)
λ: Choosiness Relaxation: $\lambda = (\mathrm{D} - \mathrm{d})/D$
D: Maximum days (Kalick and Hamilton assumed 51)
d: Current day number

Todd et al. have generated a number of papers on ABM simulation of mate choice (Todd et al. 2005; Simão and Todd 2003; Simão and Todd 2002) and built on previous work by adding an adolescent phase (where agents estimate their own quality) and a courtship phase (where agents can switch to a better partner if one becomes available). Additional design features of importance are a one-dimensional mate quality metric, an internal quality assessment that changes with interactions, and a constraint on the agent's time (i.e., the agent can only be on one date at a time). Samldino and Schank (2012) developed an ABM that builds on these previous models and considers the spatial proximity constraint on finding potential mates. Additionally, they show that these models easily fit the available data due to the complexity of the models and the limited data available.

Three papers that were found in our literature search simulate mate choice in the animal kingdom, which tangentially inform modeling online dating apps. A paper by Roswell and Cade (1993) developed a simulation to study the importance of population density, sex ratio, and the frequency of male behavior in maintaining alternative forms of behavior in a population of field crickets. The simulation most resembles an ABM. A paper by Wachtmeister and Enquist (2000) use a model to produce a hypothesis for courtship rituals in monogamous pairs. Though they do not refer to it as such, the model resembles an ABM. The model contains interacting agents that evolve over generations where the female agents containing a Neural Network to model their behavior to accept courtship. Finally, a paper by Lessells (2005) uses a combination of stochastic dynamic programming and game theory to determine optimal mating strategies. Similar to the intent of this study, Lessells varied underlying model assumptions and behaviors to determine how the optimal strategies change.

In addition, the work in the literature on matching algorithms are also relevant. Two NetLogo models, developed by Shiba, were found in the Computational Model Library; one on symmetric two-sided matching (Shiba 2013c); the other on asymmetric two-sided matching (Shiba 2013a; Shiba 2013b).

The majority of the models found in the literature either explicitly or implicitly use the ABM paradigm. Other notable features of their models include: probability functions for mating that tend towards unity with time, a single dimensional value of attractiveness, a self-assessment of attractiveness, a stable population of agents, and balanced sex ratios. Learning from the literature, our model will include these features.

4.2 Conceptual Model

The online dating app was modeled with an ABM, as it is best suited for simulating this system due to its autonomous agents that learn, adapt, and interact with one another (Macal and North 2007). Additionally, this is the most common paradigm found in the literature. We chose to use NetLogo, due to our familiarity with the tool; however, there are many ABM tools available (e.g., Repast 3, Swarm, AnyLogic, Mesa).

The online dating model must represent several rules and behaviors of the real-world system. To define what needs to be modeled, we systematically stepped through the system by following the actions of the user to identify key rules and behaviors. First, the user enters the app. We identified three methods for representation: the simulation can maintain a constant number of users, the entrance rate is modeled with a time-invariant distribution, or the entrance rate is represented with a time-dependent distribution. Since we do not know which distribution represents reality, we modeled a constant number of users; as users leave either by pairing off, being kicked-off, or through frustration, they are replaced with new users. This is similar to other models we found in the literature.

After users enter the app, they view a set of profiles that are potential dating prospects. Human sexuality is complex and to simplify the problem, in our example we only considered heterosexual courting. We refer to this set of profiles as a as a ‘pool’. This pool is constructed by the online dating algorithm. There were two design decisions made here: pool size and modeling the algorithm. For pool size, the options are all users of the opposite sex on the app or a subset of the total. We selected to model the pool as a subset of the population in order to investigate the effect of pool size on user satisfaction.

The second modeling decision was how to model an algorithm to select the pool for each person. We identified three representations of the algorithm: the pool can be generated at random, the pool can be generated based on similarity to a single trait (e.g., attractiveness), or the pool can be generated based on similarity to several traits (e.g., location, attractiveness, education level, hobbies, user preferences). Because the intent of this paper is to investigate models of human behavior, we kept the algorithm simple. We modeled the pool generation as a random assignment.

Though not all dating apps include a stage for matching (i.e., a process that allows messaging once both parties indicate interest), we decided to include this mechanism. To model the users deciding to ‘like’ a user in their pool, we considered the factors affecting this decision. We identified that attractiveness of a potential match, the attractiveness of the user, and the number of ongoing interactions all affect the decision of the user to ‘like’ the potential match. We modeled this behavior in four ways in order to address the central hypothesis of this paper. The first three models are based on the work by Kalick and Hamilton as shown in equations (3), (4), and (5). We modified the model to include a lower cutoff on the attractiveness of the potential match. The final model altered (5) to include a workload constraint. This workload constraint is a function of messages the users sent and number of dates the user has attended (6). As the user sends messages and attends dates, the probability that they will like a user decreases, where if either the maximum messages or dates are met, the probability drops to zero.

$$PL_1 = \begin{cases} \left(\frac{A_j}{k}\right)^{C\lambda} & If\ A_j \geq A_i - \Delta A \\ 0 & Otherwise \end{cases} \tag{3}$$

$$PL_2 = \begin{cases} \left(\frac{k-|A_i-A_j|}{k}\right)^{C\lambda} & If\ A_j \geq A_i - \Delta A \\ 0 & Otherwise \end{cases} \tag{4}$$

$$PL_3 = \frac{PL_1+PL_2}{2} \tag{5}$$

$$PL_4 = \alpha \times PL_3 \tag{6}$$

Such that:
PL_q: The probability to like user j using probability model q.
ΔA: The cutoff of attractiveness below self-perception of attractiveness
λ: Choosiness Relaxation $= (T_D - T_d)/T_D$
T_D: Maximum days since last date
T_d: Days since last date
α: Workload constraint $= [(M - \mathrm{m})/M] \times [(D - \mathrm{d})/D]$
M: Maximum number of messages one can send in a week
m: number of messages sent in past week
D: Maximum number of dates one can attend in a week
d: Number of dates attended in past week

Once users have made their decision on their dating pool, the app will compute the matching. When users are matched, they are allowed to message each other through the app. The decision to message is similar to the pool evaluation in its dependency on each party's attractiveness, choosiness, and their current workload. However, messaging is complicated because of several factors; here we identify two. First, culturally, it may be more common for one sex to pursue and message more. Second, the time between messaging matters, because a user may not want to message too soon and seem desperate, or message too late and miss an opportunity. Like our modeling of dating pool evaluation, we model the messaging behavior in several ways to support the hypothesis of this paper. Equation (7) shows the probability of sending a message as a function of each user's attractiveness, time since last date, workload constraints, and cultural impact. It is based on the fourth PL equation (6) with an additional factor, β, that takes a value of 0.5 for the sex that messages more and a 1.0 for the other sex. Equation (8) shows the second model for messaging where the time since the last received message is considered. Here we assume there is an optimal time to send a response, T_M, and after a while the probability to send a message goes to zero. The third messaging model combines the effects of the first two (9).

$$PM_1 = PL_4^{\beta} \tag{7}$$

$$PM_2 = \gamma \times PL_4 \tag{8}$$

$$PM_3 = \gamma \times PL_4^{\beta} \tag{9}$$

Such that:
β: Cultural effect on messaging (1 for one sex, 0.5 for the other)
γ: Effect of time since message received $\gamma = 1 - |T_M - T_m|/T_M$
T_M: The preferred time to send a follow up message
T_m: The time since the message was received or match was made

If messaging is successful, then users may decide to go on a date. We identified two models for this behavior. Dates can be modeled as occurring after a fixed or random number of messages. Alternatively, dates can be modeled as a function of the messages' attributes (e.g., quality, timeliness). We decided to model the decision to date as occurring after a fixed number of successful messages. Additionally, we imposed a constraint that the user can only go on one date a day. The users will then meet on the first available opening.

Whether another date is held can be modeled as a simple distribution or as a function of other factors (e.g., attractiveness). We modeled this as a distribution that is a function of each user's attractiveness and opportunities. This model is similar to Eq4 where each agent runs the function to get a probability threshold value. A random roll is then taken for each agent to see if they go on another date. If both are successful, then the next date will be scheduled as early as possible but no sooner than two days. We intrinsically determined that a two-day turnaround for another date is the most realistic. (10) shows this model, where $max(A_k)$ is the maximum attractiveness of the users this user is messaging or dating.

$$PD_1 = \left(\frac{A_j}{max(A_k)}\right)^{c} \tag{10}$$

Finally, dating users must decide to end their courtship or exit the app as a committed match. This decision can be modeled as a fixed or randomly set number of dates or as a function of several factors (e.g., the quality of the dates, their other prospects). We decided to model it as a fixed number of dates.

If a user does not find a committed match, then the user can leave the app individually. We modeled this as If-Then conditionals and a probability model. First, users do not consider leaving the app until they have been on the app for a set time, T_{tenure}. Next, if the user has a date planned or has had a date in the time defined by T_D, then the user stays on the app. Otherwise, the user leaves based on a probability that is the function of the messages they have received in the past T_D days, defined as $P_{Leave} = (m_r/T_D)$, where m_r is the number of messages received in T_D days. Therefore, if the user has received a message every day, they will stay; however, if the user received no messages they will certainly leave.

The above describes our conceptual model of the dating app and considers a small number of the design possibilities for this example problem. The dating app can be modeled in further minute detail or encompassing broader impacts. Additionally, we have defined many input variables whose value will define the model response. As illustrated, the number of design options available to the modeler is immense.

Our presented model and discussion does not claim to be the unerring model, but instead helps us highlight the effect of different behavior models in CAS. The design flexibility from input values is recognized by the modeling community, where sensitivity analysis on the inputs is commonly performed. This flexibility makes it easy to fit small datasets, an issue addressed by the community (Smaldino and Schank 2012). The structural uncertainty of the model (e.g., which behaviors are modeled, how behaviors are modeled) receives less attention; however, it may have a larger effect on the output measures.

5 VERIFICATION AND VALIDATION

This model was verified primarily using the tracing method. Throughout the programmed model, print statements were added, which enabled us to check that the calculations and methods functioned as intended. Each equation defined in the Conceptual Model section was hand-calculated to assure accuracy. Additionally, we utilized the NetLogo inspect function, which allows the modeler to observe the agent values at each time step. Using these methods, we have reasonable confidence that the programmed model was built as intended.

We had limited information in order to validate the programmed model using objective methods. Informed by discussion in the literature, we can expect to see a positive correlation in attractiveness between the couples. Since we have defined four pool evaluation behaviors and three messaging behaviors, we have twelve models to validate. We observed positive correlation for each of the models. We found two interesting results: the correlation ranged from 0.08 to 0.93 with all but one having a correlation below 0.55, and we observed in eleven models that the vast majority of couples formed had attractiveness values greater than 8. Though dating occurred at all levels, albeit reduced for lower attractiveness, only those in the higher end of attractiveness decided to commit to a partner. The one model that showed different results used PL_4 and PM_1. This model showed a large number of couples formed and a correlation of attractiveness of 0.93 across the attractiveness range. Without an extensive objective dataset, it is impossible to argue which aligns most with reality. We failed to invalidate any of the twelve models.

6 ANALYSIS AND RESULTS

Each model ran for 90 simulation days. To address the design hypothesis (i.e., that larger pool size leads to less satisfied customers) pool sizes of 5 through 25 at steps of 5 were run. Each run was replicated 40 times to estimate the mean number of couples formed and number of users that left without coupling. Our objective is to have more couples formed and fewer users that leave.

Figure 1 shows the results for the mean couples formed with 95% confidence intervals for increasing pool size. A linear regression of the couples formed to the pool size is shown with a red line. Each column shows results for models using the same probability to message model (i.e., PM). Each row shows results for models using the same pool evaluation model (i.e., PL).

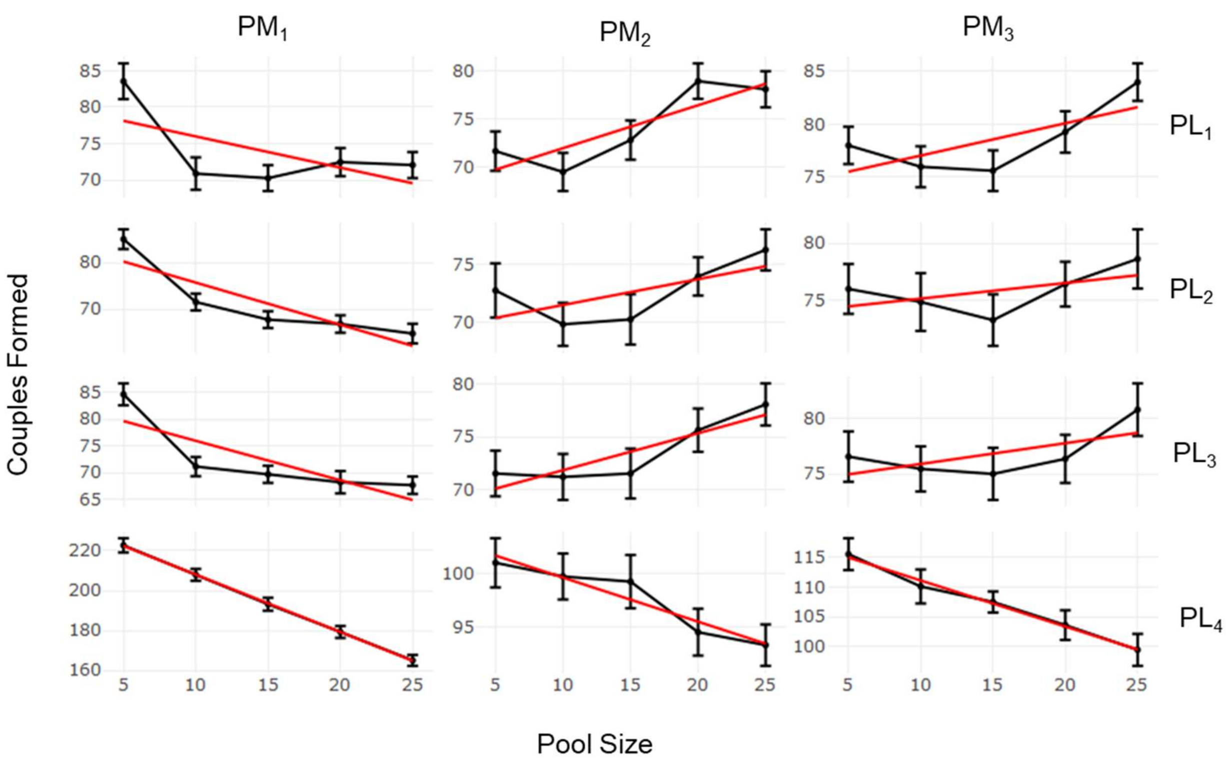

Figure 1: Results for the mean couples formed with 95% confidence intervals and linear regression for increasing pool size.

We found that the type of behavior model used drives our conclusion on whether or not smaller pool sizes are beneficial. We consider a statically significant fit to be one with a p-value less than 0.01. A summary of regression fits is shown in Table 1. The only model found to not have a significant relationship to Pool Size is model PL_2PM_3. The remaining models indicate both positive and negative relationship between couples formed and Pool Size. If we consider the models that use either PL_1, PL_2 or PL_3 with models PM_2 or PM_3, then the data suggests that a larger pool size leads to more couples formed. Conversely, models that use PL_4 or PM_3 suggests that larger pool sizes lead to fewer couples formed. Based on this data we were unable to evaluate our design hypothesis because the behavior models suggest various conclusions.

Table 1: Couples formed model fit summary.

Model	Intercept	Coeff	p-value	R^2
PL1PM1	80.26	-0.43	0.0000	0.14
PL1PM2	67.48	0.45	0.0000	0.20
PL1PM3	73.96	0.31	0.0000	0.11
PL2PM1	84.78	-0.91	0.0000	0.46
PL2PM2	69.30	0.22	0.0006	0.06
PL2PM3	73.75	0.14	0.0647	0.02
PL3PM1	83.27	-0.74	0.0000	0.38
PL3PM2	68.32	0.35	0.0000	0.12
PL3PM3	74.03	0.19	0.0093	0.03
PL4PM1	236.68	-2.87	0.0000	0.82
PL4PM2	103.72	-0.41	0.0000	0.15
PL4PM3	118.87	-0.78	0.0000	0.33

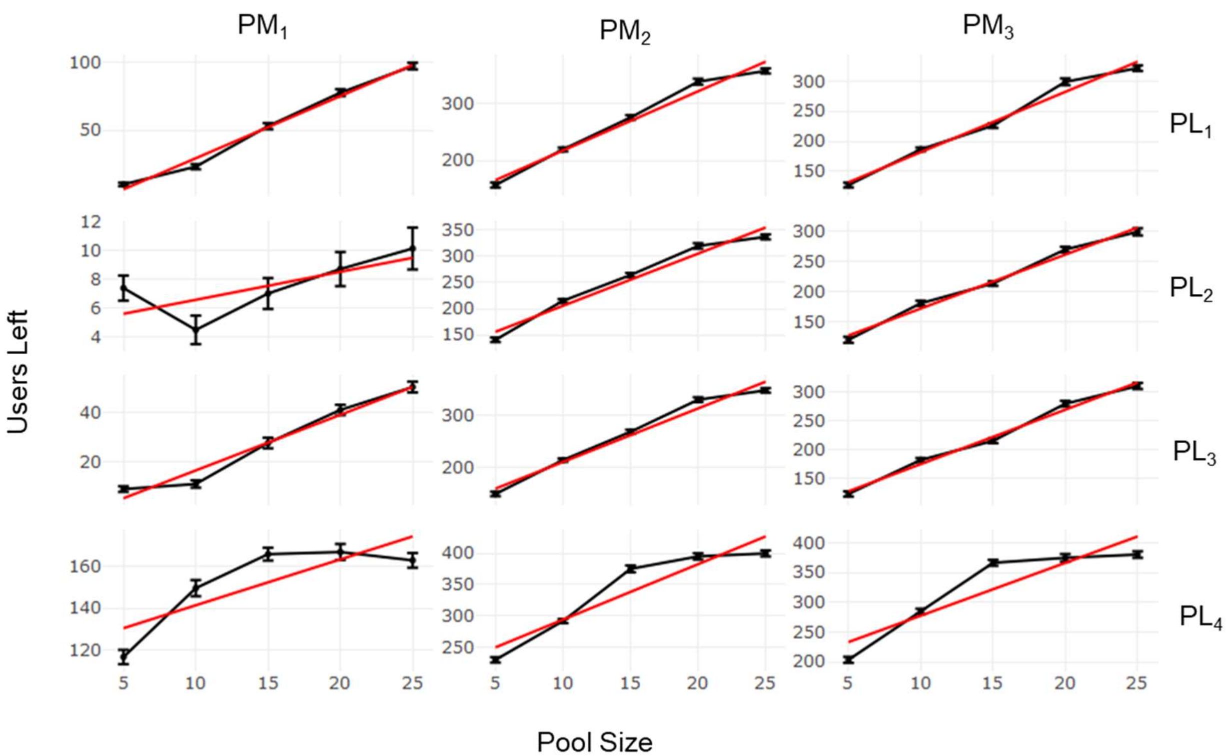

Figure 2: Results for the mean users left with 95% confidence intervals and linear regression for increasing pool size.

Figure 2 shows the number of users that left the app without finding a committed match. The effect of pool size is much more pronounced here. All behavior models tend to agree that larger pool sizes lead to more people leaving the app. The summary of the linear regression fits is shown in Table 2. These results support our design hypothesis, that larger pool size leads to greater dissatisfaction.

Table 2: Users left model fit summary.

Model	Intercept	Coeff	p-value	R^2
PL1PM1	-16.12	4.56	0.0000	0.95
PL1PM2	112.45	10.50	0.0000	0.95
PL1PM3	81.37	10.05	0.0000	0.95
PL2PM1	4.62	0.19	0.0000	0.12
PL2PM2	106.54	9.91	0.0000	0.93
PL2PM3	81.98	8.94	0.0000	0.94
PL3PM1	-6.43	2.28	0.0000	0.86
PL3PM2	106.11	10.38	0.0000	0.95
PL3PM3	81.58	9.37	0.0000	0.95
PL4PM1	119.54	2.19	0.0000	0.50
PL4PM2	205.24	8.87	0.0000	0.84
PL4PM3	188.48	8.88	0.0000	0.79

The above results support this paper's overall hypothesis that the uncertainty in behavioral modeling can have an effect on the results and conclusion of the study. The above results show both that small pool sizes and that large pool sizes can have a detrimental effect on the number of couples formed. Without sufficient objective data for validation, investigation of the simulation results by itself is insufficient to provide guidance on the validity of the model. For many CAS it is unlikely we will have sufficient objective data due to the common non-observability of the system. Therefore, the community should work together toward identifying the best behavior models in order to inform future model developments.

7 CONCLUSIONS

This paper investigated the impact of behavioral modeling assumptions for a CAS model. Behavioral modeling presents a challenge to modelers – particularly due to the difficulty in understanding and modeling human behavior, general constraints on cost and time for model development, and limited options for validation of the model. These challenges can lead to overconfidence in predictions. By modeling an example CAS with different representations of its behavior, we highlighted the impact of behavioral modeling on system design and the challenges CAS modelers face.

Our example problem was to design a dating app that provides its users with an overall satisfying experience. Our design hypothesis is that an app that presents too many options to a user will lead to a lower user satisfaction. This hypothesis was partially supported by our simulation results. The twelve models analyzed consistently showed a positive relationship between the number of users that left the simulation without a match and increasing pool size. This partially support the design hypothesis for smaller pool sizes. However, the number of couples formed for increasing pool size was found vary significantly based on the behavioral models chosen. The different models provided positive, negative, and no relationship between pool size and the number of couples formed. Though there is an indication that larger pool sizes lead to more people leaving without a partner, we cannot conclude anything on the effect on couples formed.

The observed effect of the behavior models on the number of couples formed demonstrates the sensitivity of the simulation results to the behavioral model used. Each behavioral model was similar to the others with additional terms added or removed. The models had similar structures in their mathematical form and each were probability models. We did not compare different behavioral modeling paradigms (e.g., if-then conditionals, probability models, behavior trees, neural nets). Despite this similarity, our results demonstrated that some behavioral models show a positive relationship between number of couples formed and the pool size, while other behavior models showed the converse.

It should not surprise any modeler that different models produce different results. This is why objective validation to real world data is a required step in the simulation life cycle. CAS present us with a challenge because real world data is rarely available. Without sufficient data for validation, investigation of the simulation results by itself is often insufficient to provide guidance on the validity of the model. A modeler can address this issue by investigating alternative behavioral models to identify robust trends; however, time and resource constrains often prevent the modeler from doing so. Therefore, the modeler should rely on their community for best practices and recommendations, where the community has dedicated research that compares and analyzes behavioral models of CAS phenomena.

We believe the field of CAS is best positioned to take the lead on addressing these behavioral modeling challenges because of our experience with, and core dependency on, behavioral models. We encourage the modeling and simulation community to undertake a number of actions to continue advancing the science of behavioral modeling. First, when behavioral models are presented without sufficient objective validation data, one should expect to see several conceptual models included for analysis to determine the robustness of observed emergent behaviors and quantitative results. Second, experimental replication and reporting on models has been a difficult task and under performed for decades. More replication and analysis of behavior models is a critical step for our community. Third, a meta-analysis type approach is needed for behavior models. Here, a scientist can replicate multiple proposed models from the literature on the same phenomena to investigate the different emergent properties they exhibit. They may then provide guidance to future modelers on the best conceptual models to implement. Finally, we should provide new modelers with our summarized conclusions and guidance for modeling our investigated behaviors. This will provide them with the resources they need when faced with limited resources and time. Through this open comparison and discussion, the community can gain improved understanding of CAS, better confidence in our results, and advance the scientific art of behavioral modeling.

REFERENCES

Barabási, A. and R. Albert. 1999. "Emergence of Scaling in Random Networks". *Science* 286(5439):509-512.

Borshchev, A. and A. Filippov. 2004. "From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools". In *Proceedings of the 22nd International Conference of the System Dynamics Society*, July 25th-29th, Oxford, England.

Bouloiz, H., E. Garbolino, M. Tkiouat, and F. Guarnieri. 2013. "A System Dynamics Model for Behavioral Analysis of Safety Conditions in a Chemical Storage Unit". *Safety Science* 58:32-40.

Brailsford, S. C., V. A. Lattimer, P. Tarnaras, and J. C. Turnbull. 2004. "Emergency and On-Demand Health Care: Modelling a Large Complex System". *Journal of the Operational Research Society* 55:34-42.

Centola, D. 2010. "The Spread of Behavior in an Online Social Network Experiment". *Science* 329(5996):1194-1197.

Cooke, D. L. 2003. "A System Dynamics Analysis of the Westray Mine Disaster." *System Dynamics Review* 19(2):139-166.

Folcik, V. and C. G. Orosz. 2006. "An Agent-Based Model Demonstrates that the Immune System Behaves Like a Complex System and a Scale-Free Network". In *Tenth Annual Swarm Agent-Based Simulation Meeting SwarmFest*, June 23rd-24th, University of Notre Dame, Indiana, USA.

Groff, E. R., S. D. Johnson, and A. Thornton. 2018. "State of the Art in Agent-Based Modeling of Urban Crime: An Overview". *Journal of Quantitative Criminology* 35(1):1-39.

Guimera, R., M. Sales-Pardo, and L. A. N. Amaral. 2004. "Modularity from Fluctuations in Random Graphs and Complex Networks". *Physical Review E* 70(2):025101.

Hausken, K. and J. Zhuang. 2011. "Defending Against a Terrorist Who Accumulates Resources". *Military Operations Research* 16(1):21-39.

Henrich, J., S. J. Heine, and A. Norenzayan. 2010. "The Weirdest People in the World?". *Behavioral and Brain Sciences* 33(2-3):61-83.

Kalick, S. M. and T. E. Hamilton. 1986. "The Matching Hypothesis Reexamined". *Journal of Personality and Social Psychology* 51(4):673-682.

Lessells, C. M. 2005. "Why Are Males Bad for Females? Models for the Evolution of Damaging Male Mating Behavior". *The American Naturalist* 165(S5):S46-S63.

Lucas, T. W., S. M. Sanchez, T. M. Cioppa, and A. I. Ipekci. 2003. "Generating Hypothesis on Fighting the Global War on Terrorism". In *Maneuver Warfare Science 2003,* edited by G. Horne and S. Johnson, 107-124, Monterey, CA, USA: Naval Postgraduate School Department of Operations Research.
Macal, C. and M. J. North. 2010. "Tutorial on Agent-Based Modelling and Simulation". *Journal of Simulation* 4(3):151-162.
Macal, C. M. and M. J. North. 2007. "Agent-Based Modeling and Simulation: Desktop ABMS". In *Proceedings of the Winter Simulation Conference*, edited by S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, 95-106. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
Mitchell, M. 2009. *Complexity: A guided tour.* Oxford, New York, USA: Oxford University Press.
Pan, X., C. S. Han, K. Dauber, and K. H. Law. 2007. "A Multi-Agent Based Framework for the Simulation of Human and Social Behaviors During Emergency Evacuations". *AI & Society* 22(2):113-132.
Pashler, H. and E.-J. Wagenmakers. 2012. "Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence?". *Perspectives on Psychological Science* 7(6):528-530.
Raczynski, S. 2004. "Simulation of the Dynamic Interactions Between Terror and Anti-Terror Organizational Structures". *Journal of Artificial Societies and Social Simulation* 7(2).
Rehman, T. 2016. "Historical Context of Behavioral Economics". *Intellectual Economics* 10(2):128-132.
Reynolds, C. 1995. "Boids Background and Update". https://www.red3d.com/cwr/boids/, accessed 19th March 2019.
Rowell, G. A. and W. H. Cade. 1993. "Simulation of Alternative Male Reproductive Behavior: Calling and Satellite Behavior in Field Crickets". *Ecological Modelling* 65(3-4):265-280.
Sargent, R. G. 2013. "Verification and Validation of Simulation Models". *Journal of Simulation* 7(1):12-24.
Schelling, T. C. 1978. "Micromotives and Macrobehavior." New York, New York, USA: W. W. Norton.
Schwartz, B. 2004. "The Paradox of Choice: Why More is Less". New York, New York, USA: HarperCollins.
Shiba, N. 2013a. "Analysis of Asymmetric Two-Sided Matching: Agent-Based Simulation with Theorem-Proof Approach". *Journal of Artificial Societies and Social Simulation* 16(3).
Shiba, N. 2013b. "Asymmetric Two-Sided Matching (Version 1.1.0)". CoMSES Computational Model Library https://www.openabm.org/model/3577/version/2, accessed 19th March 2019.
Shiba, N. 2013c. "Symmetric Two-Sided Matching. (Version 1.1.0)". CoMSES Computational Model Library https://www.comses.net/codebases/3575/releases/1.1.0/, accessed 19th March 2019.
Simão, J. and P. M. Todd. 2003. "Emergent Patterns of Mate Choice in Human Populations". *Artificial Life* 9(4):403-417.
Simão, J. and P. M. Todd. 2002. "Modeling Mate Choice in Monogamous Mating Systems with Courtship." *Adaptive Behavior* 10(2):113-136.
Smaldino, P. E. and J. C. Schank. 2012. "Human Mate Choice is a Complex System". *Complexity* 17(5):11-22.
Smith, A. and M. Anderson. 2014. "5 Facts About Online Dating." Pew Research Center, http://tony-silva.com/eslefl/miscstudent/downloadpagearticles/onlinedating-pew.pdf, accessed 19th March 2019.
Sterman, J. D. 1992. "System Dynamics Modeling for Project Management". *Unpublished Manuscript*, Sloan School of Management, Massachusetts Institute of Technology, Cambridge, MA, USA. http://web.mit.edu/jsterman/www/SDG/project.pdf, accessed 19th March 2019.
Szabo, C. and Y. M. Teo, and G. K. Chengleput, 2014. "Understanging Complex Systems: Using Interaction as a Measure of Emergence". In *Proceedings of the Winter Simulation Conference*, edited by A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, 207-218. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
Todd, P. M., F. C. Billari, and J. Simão. 2005. "Aggregate Age-at-Marriage Patterns from Individual Mate-Search Heuristics". *Demography* 42(3):559-574.
Turner, A., S. Balestrini-Robinson, and D. Mavris. 2012. "Analysis of the Impact of Behavioral Assumptions in Agent Based Modeling." *80th MORS Symposium*, June 11th-14th, Colorado Springs, Colorado, USA.
Wachtmeister, C.-A. and M. Enquist. 2000. "The Evolution of Courtship Rituals in Monogamous Species." *Behavioral Ecology* 11(4):405-410.

AUTHOR BIOGRAPHIES

Andrew Turner received his PhD, MS, and BS in Aerospace Engineering from the Georgia Institute of Technology. His graduate research focused on the development and validation of models of non-observable systems. Afterwards, Dr. Turner worked for MITRE in the Systems Engineering Technical Center where he focused on modeling systems and system of systems to support various US Government projects. He currently works at the Northrop Grumman Corporation. He currently serves on the board of directors of the Military Operations Research Society. His research interests include topics on the modeling, validation, and analysis of systems and system of systems. His email address is andrew.turner22@gmail.com.

Jennifer Kuczynski is a Senior Systems Engineer at MITRE. She received her M.S. in Systems Engineering and B.S. in Physics from the George Washington University. She has research experience in the areas of systems engineering, biophysics and nuclear physics. Her email address is jkuczynski@mitre.org.