
Proceedings Paper:

Vu, T.M. orcid.org/0000-0002-2540-8825, Davies, E., Buckley, C. et al. (2 more authors) (2021) Using multi-objective grammar-based genetic programming to integrate multiple social theories in agent-based modeling. In: Ishibuchi, H., Zhang, Q., Cheng, R., Li, K., Li, H., Wang, H. and Zhou, A., (eds.) Evolutionary Multi-Criterion Optimization : EMO 2021 Proceedings. EMO 2021: 11th International Conference on Evolutionary Multi- Criterion Optimization, 28-31 Mar 2021, Shenzhen, China. Lecture Notes in Computer Science, 12654 . Springer , pp. 721-733. ISBN 9783030720612

https://doi.org/10.1007/978-3-030-72062-9_57

This is a post-peer-review, pre-copyedit version of a paper published in Ishibuchi H. et al. (eds) Evolutionary Multi-Criterion Optimization. EMO 2021. Lecture Notes in Computer Science, vol 12654. The final authenticated version is available online at: http://dx.doi.org/10.1007/978-3-030-72062-9_57.

Using Multi-objective Grammar-based Genetic Programming to Integrate Multiple Social Theories in Agent-based Modeling

Tuong Manh Vu[0000−0002−2540−8825], Eli Davies,
Charlotte Buckley[0000−0002−8430−0347], Alan Brennan[0000−0002−1025−312X], and
Robin C. Purshouse[0000−0001−5880−1925]

University of Sheffield, Sheffield, UK
`t.vu@sheffield.ac.uk`, `eliwgdavies@gmail.com`,
`{c.m.buckley,a.brennan,r.purshouse}@sheffield.ac.uk`

Abstract. Different theoretical mechanisms have been proposed for explaining complex social phenomena. For example, explanations for observed trends in population alcohol use have been postulated based on norm theory, role theory, and others. Many mechanism-based models of phenomena attempt to translate a single theory into a simulation model. However, single theories often only represent a partial explanation for the phenomenon. The potential of integrating theories together, computationally, represents a promising way of improving the explanatory capability of generative social science. This paper presents a framework for such integrative model discovery, based on multi-objective grammar-based genetic programming (MOGGP). The framework is demonstrated using two separate theory-driven models of alcohol use dynamics based on norm theory and role theory. The proposed integration considers how the sequence of decisions to consume the next drink in a drinking occasion may be influenced by factors from the different theories. A new grammar is constructed based on this integration. Results of the MOGGP model discovery process find new hybrid models that outperform the existing single-theory models and the baseline hybrid model. Future work should consider and further refine the role of domain experts in defining the meaningfulness of models identified by MOGGP.

Keywords: Inverse generative social science · Agent-based modeling · Multi-objective optimization · Grammar-based genetic programming

1 Introduction

1.1 Background

Agent-based modeling (ABM) is a bottom-up methodology that models a system as a collection of heterogeneous agents and their interactions. Since ABM allows the modeling of a complex system at a fine-grained resolution, it has become an established tool for generative social science: how micro-level agent behaviours and their interactions can generate macro-level social phenomenon. Following the motto of generative social science – "If you didn't grow it, you didn't explain it." [7] – the *forward* approach is where, if an agent-based model with a defined set of mechanisms can produce the target social phenomenon, the model is a candidate explanation for the phenomenon. However, this does not mean that the candidate is unique – there can be other models that can generate the same target phenomenon. In the *inverse* problem, from a target phenomenon, the aim is to find possible explanatory agent-based models. This process is known as *inverse generative social science* [18] or *model discovery* [10]. The literature on model discovery is very limited and was reviewed by Vu and colleagues [17].

A social phenomenon can be explained by different theories. An agent-based model encoded with mechanisms from a single social theory can be a candidate explanation. But one can wonder which theory is better or what can be missing from a theory. There is also a possibility that a single theory cannot explain the phenomenon and there is a need to combine multiple theories. However, the integration of multiples theories is a major endeavour. To address that gap, this paper proposes using genetic programming to explore combinations of social theories with the aim of better explanatory capability.

In genetic programming [14], a population of computer programs or candidates is evolved over many generations using a fitness function. Computer programs with high fitness are selected probabilistically for crossover and mutation operators to generate the next generation. To mitigate the problem of invalid computer programs created by the random nature of the genetic operators, we adopted grammar-based genetic programming (GGP) [12] in which a grammar is employed for genotype-to-phenotype mapping to enforce a particular structure and guide the evolutionary process. In this paper, the GGP was used in the model discovery process to explore the search space of multi-theory models.

1.2 Aim of the present study and organization of the paper

This paper aims to demonstrate how to use multi-objective GGP (MOGGP) to integrate single-theory models into a hybrid model to provide better explanation of a social phenomenon. As a case study, the paper uses two established agent-based models based on social norm theory and social role theory for the integration and model discovery process. In Section 2, the integration and MOGGP model discovery processes are both introduced. Section 3 briefly introduces the existing models that form the basis for integration. Results and discussion of the MOGGP process are provided in Section 4. Lastly, Section 5 concludes the paper and suggests possible future works.

2 Model discovery process

The flow chart depicting the model discovery process is shown in Figure 1. This is an extension of the recent approach by Vu and colleagues [17]. There are three roles in the model discovery process: *domain expert*, *analyst*, and *modeler*.

The domain expert is a person with deep knowledge about the social science underpinning the model and who can assess the theoretical credibility of the model structure. The analyst assesses commonalities and differences between social theories with the domain expert, as well as abstracting and defining the building blocks of entities and mechanisms. The modeler designs, implements, verifies and validates the agent-based models.

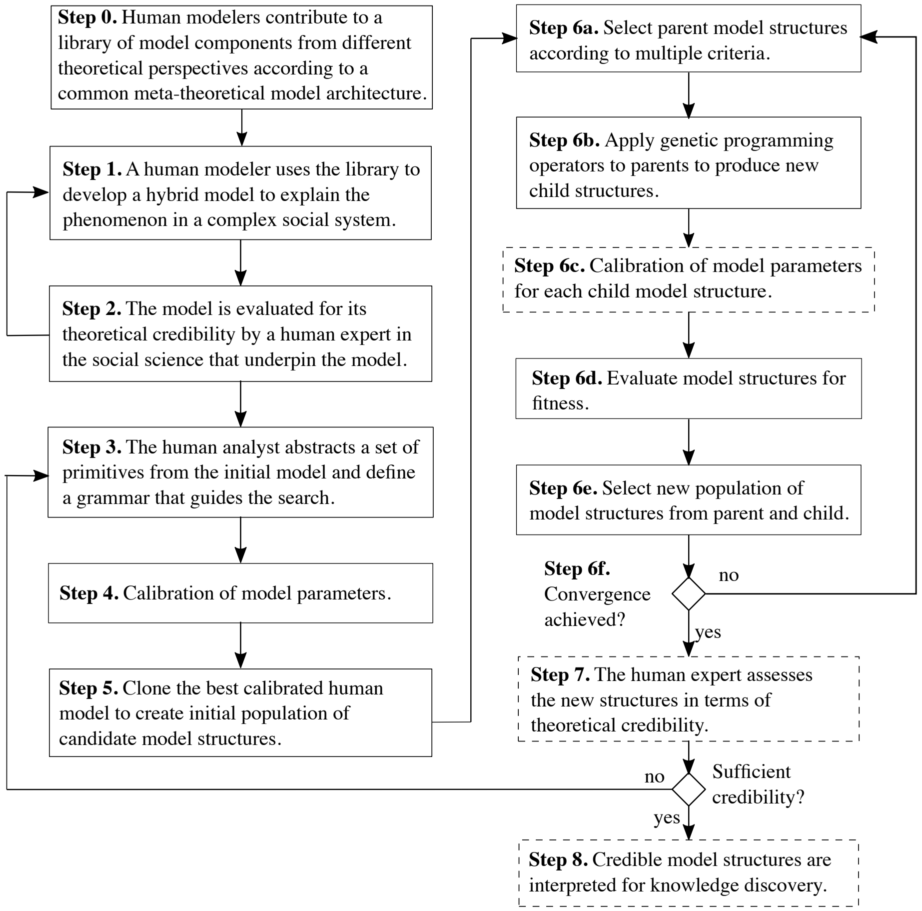

Fig. 1. Model discovery process

Step 0 represents a pre-condition for the model integration process. We assume that a library of theory building blocks, implemented as model components according to a common meta-theoretical software architecture is available [19]. In **Step 1**, a human modeler uses the library to develop an initial hybrid model. In **Step 2**, the theoretical credibility of the model is evaluated by a human expert.

The model may need to be adjusted and redeveloped (returning to Step 1). In **Step 3**, the human analyst abstracts a set of primitives from the hybrid model and defines the grammar for the genetic programming. In **Step 4**, model parameters are calibrated. In **Step 5**, the hybrid model built by the human modeler with the best calibrated setting is cloned for the whole initial population.

The evolutionary process is performed in **Step 6**. Parent candidates are selected based on fitness in Step 6a. In Step 6b, a new population of child structures is produced using genetic operators (crossover and mutation). In Step 6c, the model parameters are re-calibrated. However Step 6c was omitted due to the computational intensity of such a nested calibration approach and limits on the available computing resources. In Step 6d, all candidates are evaluated for fitness (such as model error when comparing simulated data with empirical data). A new population is selected in Step 6e. The evolutionary process is continued until a stopping condition, such as convergence or maximum iteration, is achieved.

In **Step 7**, the domain expert discusses with the modeler and the analyst about the theoretical credibility of the new models. If the new structures are not credible, the grammar can be changed (Step 3) to better guide the search or have more meaningful operations. In **Step 8**, the new structures are interpreted for knowledge discovery and theory development. Due to the limited space in this paper, we omitted steps 7 and 8 to focus on demonstrating the methodology.

3 A hybrid model of norm theory and role theory

3.1 Alcohol use modeling

In this paper, the integration process is performed on generative models of alcohol use behavior in the US population. Each model uses concepts from a theory of alcohol use and expresses them as equations to generate alcohol use behaviors in simulated individuals. We use empirical data from the Behavioral Risk Factor Surveillance System (BRFSS) [6] to generate alcohol use targets to calibrate our models. Specifically, prevalence of alcohol use in the past 12 months, average quantity of alcohol consumed per day (grams of ethanol) and average frequency of alcohol use (drinking days per month). Models are built according to a mechanism-based social systems modeling (MBSSM) software architecture, following a *general micro-macro scheme* which describes the dynamic interplay between individuals and social structures, leading to emergent population-level patterns in alcohol use [19]. Individuals in the models are from a representative population-level US microsimulation model 1980-2015 that accounts for births, deaths and migration over time [2]. Individuals in the microsimulation have socio-demographic properties and alcohol use variables at baseline that are generated from several US data sources including the BRFSS, US Census [16] and American Community Survey [13].

3.2 Social norms model

The social norms model is described in detail in [15]. This model uses concepts from social norm theory to generate a disposition to consume alcohol for each modeled individual. Disposition ($NormDisp$) is determined by three variables: descriptive norm, injunctive norm and autonomy, which all range between 0 and 1. The descriptive norm is an individual's appraisal of drinking behavior in their reference group (age group and sex) in terms of whether individuals are drinking at all (prevalence) and how much they are drinking (quantity). The descriptive norm is determined by calculating the prevalence and quantity of drinking in the reference group and applying a perception bias that adjusts the descriptive norm to be closer to the agent's own drinking. A descriptive norm of 1 would indicate that 100% of individuals in the reference group are perceived to be drinking. The injunctive norm refers to the perceived acceptability of drinking in society for an individual like them (i.e., in their reference group), initialized using data from the US National Alcohol Survey [9]. An injunctive norm of 1 would indicate that it is completely acceptable for an individual in a reference group to drink. Autonomy refers to the individual's desire to ignore the norms; an autonomy of 0.7 would indicate that the agent only pays attention to norms 30% of the time. These concepts are combined in Equation 1 to determine, for individual i, disposition $NormDisp$ to consume drink k. The injunctive norms I are combined with descriptive norms D and weighted by the inverse of autonomy a. This term is combined with the individual's desire u to consume drink k weighted by autonomy. Over time, the injunctive and descriptive norms are updated in response to individuals' drinking behavior in the model.

$$NormDisp_i[k] = u_i[k]a_i + (1 - a_i)\sqrt{D_i[k]I_i[k]} \tag{1}$$

3.3 Social roles model

The social roles model is described in detail in [17] and is based on social role theory [1] that describes how individuals' marital, parental and employment responsibilities affect their desire and ability to drink. Disposition to drink according to social role theory ($RoleDisp$) is generated using Equation 2 and is determined by an individual's desire to drink, u, for drink k multiplied by their probability of having the opportunity to drink inside ($ProbOpIn$) and outside ($ProbOpOut$) the home. Social role theory suggests that the roles individuals hold affect their ability to participate in drinking situations and therefore regulate the daily opportunities to consume alcohol [11]. Opportunity to drink in and out of the home is determined by role load (the stress that results from needing to perform a role), which is calculated using each individual's role status and their level of involvement in that role. Role strain – "the experience of stress associated with positions or expected role" [1] – is suggested to lead to alcohol use as a means of coping with a set of roles that are too complex (role load) or lacking roles that provide meaning (role deprivation). Role strain ($RoleStrain$) is weighted by β, describing the size of the effect role strain has on drinking behavior. On each day, $RoleDisp$ is calculated for each individual for $k = 1$ and compared to a uniform random number between 0 and 1 to determine if they will drink. For each individual that drinks, disposition is calculated for $k = 2$, and continues for progressive values of k until they stop drinking.

$$RoleDisp_i[k] = u_i[k](ProbOpOut_i[k] + ProbOpIn_i[k])(1 + \beta RoleStrain_i[k]) \quad (2)$$

3.4 Integration: a hybrid model of norm theory and role theory

We postulated that the social theories have different influence on an individual's k^{th} drink. For example, let us imagine an employed individual, *Agent Alan*, engages in a drinking occasion. Agent Alan's first two drinks are based solely on social norms. Agent Alan's chance of consuming a third drink is lower because he is aware he has to go to work the next day – the balance of probabilities being based on both norms and roles. Agent Alan never drinks four drinks because of his prospective employment commitments, so the disposition to consume a fourth drink is based only on roles. Extending this hypothesis, we decided to perform model discovery on the drinking disposition of individual agents.

Both the social norms and social roles models operate by calculating a disposition (*Disp*) for each agent on each day to determine whether to drink and, if so, how much to drink. In the hybrid model, the social norms and roles dispositions are calculated on each day. The models are combined by using a weighting (w, bounded between 0 and 1) of the two calculated dispositions ($NormDisp$ and *RoleDisp*) as shown in Equation 3. The hybrid model was implemented using MBSSM [19] and Repast HPC libraries [4].

$$Disp_i[k] = wRoleDisp_i[k] + (1 - w)NormDisp_i[k] \quad (3)$$

3.5 Parameter calibration

The hybrid model contains 65 parameters that represent unknown effect sizes from social role theory and social norm theory plus the weighting for combining the dispositions. Each parameter is assigned a prior distribution reflecting the range of acceptable values. We sampled 10,000 parameter settings from the joint prior distribution using the lhs R package [3]. Each model setting was run once for the years 1984–2004 and an implausibility metric (Equation 4) was calculated to compare the model output to alcohol use BRFSS target data for prevalence, quantity and frequency of male and female drinking.

$$z_1 = \frac{1}{KM} \sum_{k=1}^{K} \sum_{m=1}^{M} \frac{|y_m^\star[k] - y_m[k]|}{\sqrt{(s_m[k])^2 + (d_m)^2}}, \quad (4)$$

where M is the number of output measures, K is the number of observations, $s_m[k]$ is the observed standard error for output m at time point k, and $(d_m)^2$ is the variance of the model discrepancy [1] for output m, which is taken as 10% of the possible output range for each output.

---

[1] Model discrepancy is the error in a model output that arises because the model is not a perfect representation of reality.

3.6 Multi-objective Grammar-based Genetic Programming

In the hybrid model, an agent can have maximum of 30 drinks within a day, so k is ranged from 1 to 30. After parameter calibration, agents used the following calibrated equation to calculate all the drinking dispositions: $Disp_i[k] = 0.99667 RoleDisp_i[k] + (1 - 0.99667) NormDisp_i[k]$. For model discovery, the MOGGP can be set up to work with 30 equations for each of the k^{th} drinks. However, for simplification, we decided to use only six equations by grouping the drinks into the following categories: drink 1, drink 2, drink 3 to 4, drink 5 to 7, drink 8 to 11, drink 12+. For the primitives, terminals consist of norms disposition and role disposition and the allowed functions are weighted sum and geometric mean. Additionally, a list of both arbitrary and calibrated constants are included. The design of the MOGGP is captured in the grammar:

```
<p> ::= mediatedFirstDrink=<e>; mediated2Drinks=<e>;
  mediated3to4Drinks=<e>; mediated5to7Drinks=<e>;
  mediated8to11Drinks=<e>; mediated12MoreDrinks=<e>;
<v> ::= RoleDisp | NormDisp
<e> ::= (<c>*<e>+<c>*<e>) | sqrt(<e>*<e>) | <v>
<c> ::= 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
  0.99667 | 0.00333
```

For the initial population, we used the same structure designed and calibrated by the modeler. Then a MOGGP was configured to minimize both model error and complexity. The first objective, model error, represents the ability of the simulation model to reproduce the pattern observed in the real world and is captured by Equation 4. The second objective, complexity, is defined by the number of nodes in the structures and discourages complex structures that overfit the data and are difficult for humans to interpret. In this paper, we used the NSGA-II optimizer [5] to develop an even, sample-based representation of the Pareto front representing the trade-off between model error and complexity.

The MOGGP was implemented using the PonyGE2 toolkit [8], which supports both NSGA-II optimization and the complexity evaluation of a structure via number of nodes in the tree. Model error evaluation of a new model structure required additional scripting: (i) modifying the source code of the drinking disposition in agent decision making; (ii) re-compiling the Repast HPC model; (iii) running the simulation; and (iv) calculating the model error by comparing simulated data against empirical data.

The evolutionary setup was as follows: 500 candidates per population for 50 generations, 75% subtree crossover, 25% subtree mutation, tree depth maximum of 17, and other default settings of PonyGE2. The source code is available at https://bitbucket.org/r01cascade/emo2021_hybrid_norms_roles_gp and is licensed under the GNU General Public License version 3. It is computationally intensive to do a complete run of MOGGP; the process took 2.5 days on an Intel i9 9980XE processor with 36 cores.

4 Results and Discussion

Both model error and complexity reduced over the generations. The search converged at generation 36, showing no change to the Pareto front. Figure 2 shows the final 14 non-dominated structures found by the MOGGP, the structure of the human hybrid model, and the two single-theory models (before integration into the hybrid model). From this point onward, the structures found by MOGGP will be referred to by their complexity value, i.e. the structure with complexity 19 will be called GP19. On the Pareto front, there are six extremely complex structures but with minimal improvement on the model errors (GP231, GP235, GP247, GP259, GP283, GP287). We decided to exclude these models from the discussion because it is very challenging to interpret them. Table 1 shows the structures of the human hybrid model and the remaining eight structures discovered by MOGGP.

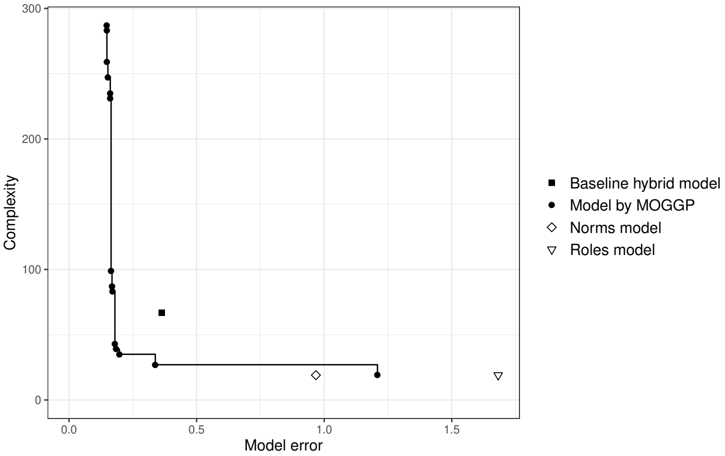

Fig. 2. Pareto front of the structures discovered by the MOGGP process versus the original baseline hybrid structure developed by the modeler

All the dispositions in the simplest model GP19 equate to $NormDisp$ which means that for a single theory, the social norms model performs better than the roles model. This is the same when comparing the two single-theory models before integration in Figure 2. Another observation is that GP19 has the same higher error even though its mechanisms are the same as the single-theory norms model. This is because GP19 used the calibrated parameters from the hybrid model. It would perform the same if we recalibrate GP19 in Step 6c. Additionally, Figure 2 shows that the two single theory models and GP19 performed poorly in term of model error when compared to the other hybrid models. Thus it can be concluded that, in this case study, integration of multiple theories better explains the phenomenon than a single theory.

Table 1. Hybrid model structures in detail, excluding six extremely complex structures. Each row of a structure relates to the ordinal drinking categories: drink 1, drink 2, drink 3 to 4, drink 5 to 7, drink 8 to 11, drink 12+.

ID	Complexity	Model error	Simplified structure of the dispositional equations
Human	67	0.3629	(0.964*RoleDisp + 0.036*NormDisp);
			(0.964*RoleDisp + 0.036*NormDisp);
			(0.964*RoleDisp + 0.036*NormDisp);
			(0.964*RoleDisp + 0.036*NormDisp);
			(0.964*RoleDisp + 0.036*NormDisp);
			(0.964*RoleDisp + 0.036*NormDisp);
GP19	19	1.2089	NormDisp;
			NormDisp;
			NormDisp;
			NormDisp;
			NormDisp;
			NormDisp;
GP27	27	0.3382	0.964*RoleDisp+0.036*NormDisp;
			RoleDisp;
			RoleDisp;
			RoleDisp;
			RoleDisp;
			RoleDisp;
GP35	35	0.1974	0.036*NormDisp + 1.7005*RoleDisp;
			RoleDisp;
			RoleDisp;
			NormDisp;
			NormDisp;
			NormDisp;
GP39	39	0.1843	0.036*NormDisp + 0.9293*RoleDisp + 0.7712*sqrt(NormDisp*RoleDisp);
			RoleDisp;
			RoleDisp;
			NormDisp;
			RoleDisp;
			NormDisp;
GP43	43	0.1799	0.036*NormDisp + 1.7005*RoleDisp;
			RoleDisp;
			NormDisp;
			0.1*NormDisp + 0.964*RoleDisp;
			RoleDisp;
			NormDisp;
GP83	83	0.1709	0.036*NormDisp + 1.83006*RoleDisp;
			0.036*NormDisp + 0.964*RoleDisp;
			0.964*NormDisp + 0.036*RoleDisp;
			1.1581*NormDisp;
			0.7*NormDisp + 0.964*RoleDisp;
			NormDisp;
GP87	87	0.1684	0.036*NormDisp + 1.71438*RoleDisp;
			0.964*RoleDisp + 0.036*sqrt(NormDisp*NormDisp);
			0.036*NormDisp + 0.964*RoleDisp;
			1.89459*NormDisp;
			0.036*NormDisp + 0.036*RoleDisp;
			RoleDisp;
GP99	99	0.1648	0.036*NormDisp + 1.71438*RoleDisp;
			0.036*NormDisp + 0.964*RoleDisp;
			0.036*NormDisp + 0.964*RoleDisp;
			1.89459*NormDisp;
			0.036*NormDisp + 0.9*RoleDisp;
			0.7712*NormDisp + 0.8072*RoleDisp;

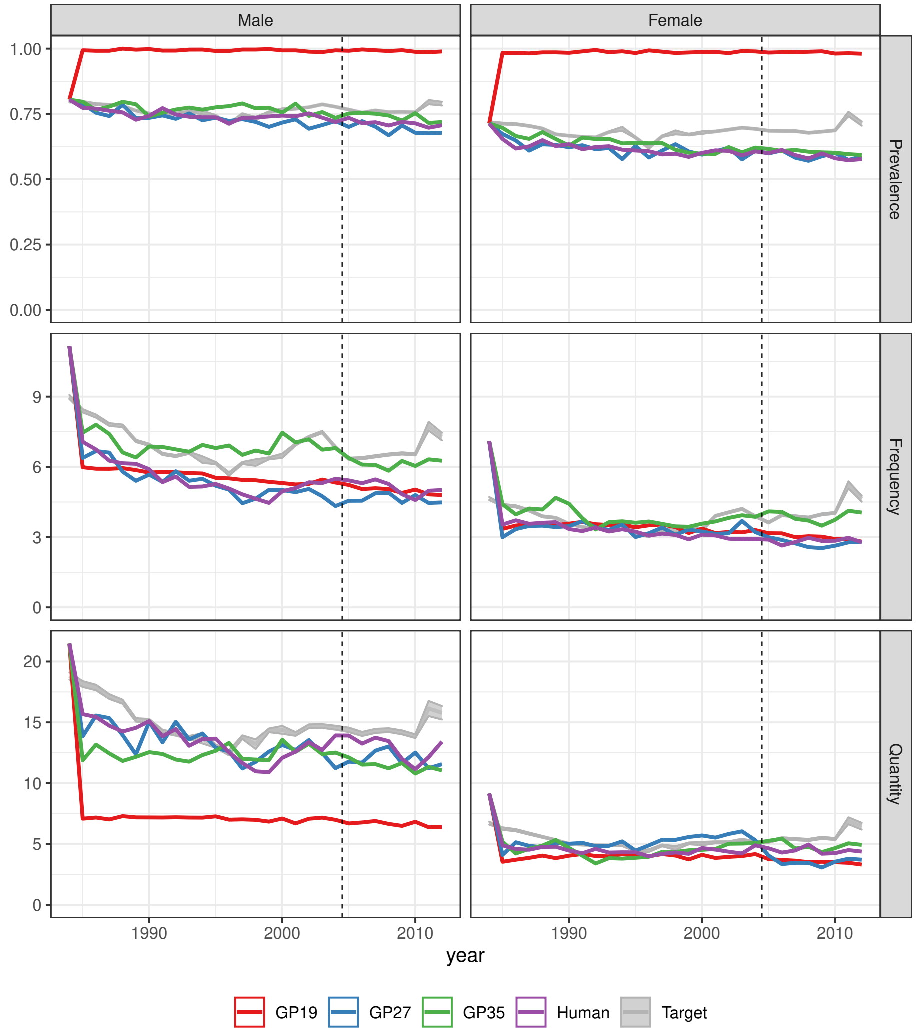

Fig. 3. Time series of the human model, GP19, GP27, and GP35 versus the empirical target data (mean and 95% confidence interval). The vertical dashed line separates the calibration period (1984–2004) and validation period (2005–2012).

Interestingly, in GP27 (the next model after GP19 in term of complexity), the first equation is the same as the baseline model, $0.964 * RoleDisp + 0.036 * NormDisp$, and the remaining equations are $RoleDisp$. This means that just by including a fraction of the norm disposition into the first drink disposition, the roles model was improved (into a hybrid model) and outperformed the norms model. The next structure is GP35 where the first three equations are dominated by role disposition and the last three equations are only influenced by norm disposition. This suggests that the first four drinks are driven by roles and the heavy drinking of 5+ drinks is driven by norms. Figure 3 shows the 1984–2012 time series of the four structures discussed (human, GP19, GP27, GP35).

For the remaining structures (GP39, GP43, GP83, GP87, GP99), the roles and norms disposition alternate between different categories. Since the domain expert asserted that there is no theory supporting this behavioral approach, we conclude that these structures are not theoretical plausible. In future work, the next iteration of the grammar should include this restriction.

5 Conclusion

This paper presents a novel method that utilizes multi-objective grammar-based genetic programming to integrate multiple social theories and discover new model structures. The case study of alcohol use modeling has shown that a multi-theory model can better explain the real world phenomenon than a single-theory model and our model discovery method offers a promising approach to generate novel combinations of multiple mechanisms of these theories. In the research frontier of inverse generative social science, our work shows the feasibility of intergating multiple social theories to better explain the targeted phenomenon in the social system. This work also highlights the challenge of *meaningful* integration, requiring the involvement of the domain expert during grammar design as well as the subsequent theoretical credibility assessment. Future research should systematically develop the role of domain experts in the discovery process.

Acknowledgment. The research reported in this paper was funded by the National Institute on Alcohol Abuse and Alcoholism of the National Institutes of Health under Award Number R01AA024443.

References

1. Biddle, B.J.: Role theory: Expectations, identities, and behaviors. Academic Press (1979)
2. Brennan, A., Buckley, C., Vu, T.M., Probst, C., Nielsen, A., Hao Bai, Thomas Broomhead, Greenfield, T.K., Kerr, W., Meier, P.S., Rehm, J., Shuper, P., Strong, M., Purshouse, R.C.: Introducing CASCADEPOP: an open-source sociodemographic simulation platform for US health policy appraisal. International Journal of Microsimulation (in press)
3. Carnell, R.: lhs R package: Latin Hypercube Samples (2019), https://github.com/bertcarnell/lhs

4. Collier, N., North, M.: Parallel agent-based simulation with Repast for High Performance Computing. SIMULATION **89**(10), 1215–1235 (Oct 2013).
5. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation **6**(2), 182–197 (2002).
6. for Disease Control and Prevention (CDC), C.: Behavioral Risk Factor Surveillance System Survey Data. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (1984)
7. Epstein, J.M.: Agent-based computational models and generative social science. Complexity **4**(5), 41–60 (1999)
8. Fenton, M., McDermott, J., Fagan, D., Forstenlechner, S., Hemberg, E., O'Neill, M.: PonyGE2: Grammatical Evolution in Python. In: Proceedings of the Genetic and Evolutionary Computation Conference Companion. pp. 1194–1201. GECCO '17, ACM, New York, NY, USA (2017).
9. Greenfield, T.K., Karriker-Jaffe, K.J., Kaplan, L.M., Kerr, W.C., Wilsnack, S.C.: Trends in Alcohol's Harms to Others (AHTO) and Co-occurrence of Family-Related AHTO: The Four US National Alcohol Surveys, 2000–2015. Substance Abuse: Research and Treatment **9**(Suppl 2), 23–31 (Oct 2015).
10. Gunaratne, C., Garibay, I.: Alternate social theory discovery using genetic programming: towards better understanding the artificial anasazi. In: GECCO 2017 - Proceedings of the 2017 Genetic and Evolutionary Computation Conference. pp. 115–122. ACM Press, Berlin, Germany (2017).
11. Knibbe, R.A., Drop, M.J., Muytjens, A.: Correlates of stages in the progression from everyday drinking to problem drinking. Social Science & Medicine **24**(5), 463–473 (1987)
12. Koza, J.: Genetic Programming: On the programming of computers by means of natural selection. MIT Press, Cambridge, MA (1992)
13. Manson, S., Schroeder, J., Van Riper, D., Ruggles, S.: IPUMS National Historical Geographic Information System: Version 14.0 [Database]. Minneapolis, MN: IPUMS. (2019)
14. Poli, R., Langdon, W.B., McPhee, N.F., Koza, J.R.: A field guide to genetic programming. Lulu Press, Morrisville, NC (2008)
15. Probst, C., Vu, T.M., Epstein, J.M., Nielsen, A.E., Buckley, C., Brennan, A., Rehm, J., Purshouse, R.C.: The normative underpinnings of population-level alcohol use: An individual-level simulation model. Health Education & Behavior (2020).
16. Ruggles, S., Flood, S., Goeken, R., Grover, J., Meyer, E., Pacas, J., Matthew Sobek: IPUMS USA: Version 9.0 [dataset]. Minneapolis, MN: IPUMS (2019).
17. Vu, T.M., Buckley, C., Bai, H., Nielsen, A., Probst, C., Brennan, A., Shuper, P., Strong, M., Purshouse, R.C.: Multiobjective genetic programming can improve the explanatory capabilities of mechanism-based models of social systems. Complexity **2020**, 1–20 (Jun 2020).
18. Vu, T.M., Probst, C., Epstein, J.M., Brennan, A., Strong, M., Purshouse, R.C.: Toward inverse generative social science using multi-objective genetic programming. In: GECCO 2019 - Proceedings of the 2019 Genetic and Evolutionary Computation Conference. pp. 1356–1363. GECCO '19, Association for Computing Machinery, Prague, Czech Republic (Jul 2019).
19. Vu, T.M., Probst, C., Nielsen, A., Bai, H., Meier, P.S., Buckley, C., Strong, M., Brennan, A., Purshouse, R.C.: A software architecture for mechanism-based social systems modelling in agent-based simulation models. Journal of Artificial Societies and Social Simulation **23**(3), 1 (2020)