This is a repository copy of *Multiobjective genetic programming can improve the explanatory capabilities of mechanism-based models of social systems*.

WILEY | Hindawi

Research Article

Multiobjective Genetic Programming Can Improve the Explanatory Capabilities of Mechanism-Based Models of Social Systems

Tuong M. Vu ©,[1] **Charlotte Buckley,**[2] **Hao Bai,**[2] **Alexandra Nielsen,**[3] **Charlotte Probst,**[4] **Alan Brennan,**[1] **Paul Shuper,**[4] **Mark Strong,**[1] **and Robin C. Purshouse**[2]

[1]*School of Health and Related Research, University of Sheffield, Sheffield, UK*
[2]*Automatic Control and Systems Engineering, University of Sheffield, Sheffield, UK*
[3]*Alcohol Research Group, Public Health Institute, Emeryville, CA, USA*
[4]*Institute for Mental Health Policy Research, Centre for Addiction and Mental Health, Toronto, Canada*

Correspondence should be addressed to Tuong M. Vu; t.vu@sheffield.ac.uk

The generative approach to social science, in which agent-based simulations (or other complex systems models) are executed to reproduce a known social phenomenon, is an important tool for realist explanation. However, a generative model, when suitably calibrated and validated using empirical data, represents just one viable candidate set of entities and mechanisms. The model only partially addresses the needs of an abductive reasoning process—specifically it does not provide insight into other viable sets of entities or mechanisms nor suggests which of these are fundamentally constitutive for the phenomenon to exist. In this paper, we propose a new model discovery framework that more fully captures the needs of realist explanation. The framework exploits the implicit ontology of an existing human-built generative model to propose and test a plurality of new candidate model structures. Genetic programming is used to automate this search process. A multiobjective approach is used, which enables multiple perspectives on the value of any particular generative model—such as goodness of fit, parsimony, and interpretability—to be represented simultaneously. We demonstrate this new framework using a complex systems modeling case study of change and stasis in societal alcohol use patterns in the US over the period 1980–2010. The framework is successful in identifying three competing explanations of these alcohol use patterns, using novel integrations of social role theory not previously considered by the human modeler. Practitioners in complex systems modeling should use model discovery to improve the explanatory utility of the generative approach to realist social science.

1. Introduction

Agent-based simulation (ABS) is a well-established tool for understanding complex systems using the generative social science approach. The goal of generative social science is to explain and understand a social phenomenon as the result of actions of autonomous entities acting according to causal mechanisms or rules as encoded in an agent-based model [1]. If a modeler encodes a model that produces a known empirical pattern, the so-called "generative test" is met, and the postulated mechanisms form a candidate explanation for the phenomenon. Many social phenomena may be explained by a multiplicity of theories, each of which could pass the generative test when encoded as an ABS, leaving us to wonder which theory is correct; how can theories be combined; and what is missing from our theories? Here, we propose a novel method of discovering new models and extending the explanatory capabilities of theory-driven generative models using multiobjective genetic programming—a process of knowledge discovery. Elements of a generative theory or several generative theories are codified in a common grammar, evolved through genetic

programming, evaluated for empirical fit, complexity, and interpretability, and interpreted by subject matter experts to bring new insight to the social phenomenon.

In this paper, we set out the following aims: (1) to explain the role of complex systems models for realist explanation; (2) to define the structural calibration method: a retroductive model discovery framework; (3) to demonstrate the application of the model discovery framework to a specific mechanism-based social systems model; and (4) to discuss the implications of computer aided model discovery in light of the case study results.

2. Methods

2.1. The Role of Abductive Reasoning in Mechanism-Based Explanation. The context for our methods is the generative [1] or mechanism-based [2] approach to the study of complex social systems. In this approach, we assume that any concrete phenomenon (which may be empirically observable to a greater or lesser degree) emerges from the dynamic interplay of *real* entities and mechanisms that exist independently of our ability to detect them [3]. In this context, the role of complex systems modeling is principally explanatory, in helping to gain insights into theorised entities and mechanisms by representing them in a dynamic simulation model.

Abductive reasoning plays a key role in mechanism-based explanations and can be conceived of in two parts [4]:

(i) Redescription: situating the concrete phenomenon as a case which emerges from the hypothesized interacting components (i.e., entities and mechanisms) of one or more theories.

(ii) Retroduction: identifying which of the components in the redescription are fundamentally constitutive to the emergence of the phenomenon (i.e., entities and mechanisms whose inexistence would preclude the phenomenon).

The development of a complex systems model by a human modeler is principally a redescription activity—the modeler uses existing theory (and potentially develops new theory) to construct a set of equations and rules that, when executed as a simulation, produce emergent outcomes that are in some sense comparable to the phenomenon under investigation. As part of the model building process, the modeler defines—either explicitly or implicitly—an ontology of real entities, the agents of agent-based simulations, and mechanisms, the rules that determine action and interaction.

The generative approach commits the modeler to at least a limited form of retroduction—the simulation, as redescription, is scrutinised for its ability to reproduce the concrete phenomenon, in so far as the latter is observable in empirical data. The simulation parameters often have to be manipulated in order to achieve good similarity; in the computational modeling community, this process is known as *calibration* [5] and is commonly identified as belonging to best practice programs for analytical sociological research [6]. If a simulation can be calibrated successfully, then the redescription it encodes is said to pass the *generative sufficiency* test—it remains a candidate explanation for the phenomenon [1]. However as a retroductive process, the generative approach, when applied to a single simulation model, has two fundamental shortcomings: (1) it does not allow the entities and mechanisms to be accepted as fundamentally constitutive since to do so would be to commit the fallacy of affirming the consequent; (2) neither does it allow the entities and mechanisms to be rejected as fundamentally constitutive, only that their present configuration or representation in the simulation model is nonconstitutive.

Together, these limitations form the basis for many of the concerns about complex systems modeling raised within the sociological community (see, for example, [7]). We argue that the limitations arise from the focus of the modeling on a single redescription, i.e., a single ABS. To improve our explanatory capability, we need to increase the number of redescriptions considered within the overall modeling process, i.e., by building multiple ABS that either interpret a single theory in different ways or represent multiple different theories or both. By subjecting a plurality of redescriptions to retroduction, we can seek to identify commonalities in the theory components that survive the generative sufficiency test; we can also seek to increase the robustness of the test outcome to potential issues with the configuration or representation of a theory within the simulation. Within the context of model calibration activities, this concept is operationalized as interrogating model structure (i.e., the selection and configuration of entities and mechanisms) in addition to model parameters—we call this *structural calibration*. Perhaps surprisingly, given its key role in the abductive reasoning process, the complex systems modeling community has yet to pay significant attention to the issue of structural calibration. Below we review the handful of existing works on this topic.

2.2. Existing Works on Structural Calibration of Simulation Models. A very small literature exists on the structural calibration (described variously as "theory discovery" [8], "model discovery" [9], and "inverse generative social science" [10]) of mechanism-based models, all of which use evolutionary computing (EC) methods to steer the search for good model structures. In the earliest known study, Smith [11] used a genetic algorithm to identify simplified representations of behavior that could reproduce the observed social assortativity of birds. Later, with a focus on reproducing observed human crowd dynamics, Zhong et al. [12] used gene expression programming to identify the structure of a reward function representing individual decision making in a "sense-think-act" framework. Gunaratne and Garibay [8] used genetic programming to revise agents' farm selection rules for the "Artificial Anasazi" model, in order to more accurately reproduce the archeological population demography of Long House Valley, Arizona. Most interestingly, again within the context of the Artificial Anasazi, the same authors then used genetic programming to identify components for Epstein's Agent_Zero model of human

behavior as the basis of farm selection decisions [9]. Finally, in work that forms a prelude to the present study, Vu et al. [10] used multiobjective genetic programming to identify alternative situational mechanisms for a social norms model of alcohol use, aimed at both improved representation of observed drinking patterns in the US over a 15-year period and theoretical interpretability (operationalized as the number of terms in the situational mechanism).

There exist a number of issues and limitations with these early techniques. While the studies have demonstrated success at improving the goodness of fit to empirical data, they have all been limited in scale—focusing on specific aspects of larger models. Further, the studies focusing on human behavior have also struggled with the issue of theoretical meaningfulness of the structures that have been identified. While minimizing or constraining the number of terms in the candidate behavioral rules is clearly helpful at improving interpretability, it is often the case that the new structures remain challenging to interpret in terms of the original theory, with this crucial activity deferred to future work.

2.3. Proposed Approach to Structural Calibration.

Here, we describe a new framework for structural calibration and position it explicitly as a tool for realist explanation that can be used alongside more traditional approaches within the realist tradition [3]. Our approach is grounded in the recognition that the human modeler uses a creative process of redescription that results in the construction of an ontology for entities and mechanisms that may be implicated in the generation of a complex phenomenon. The starting point for our framework is this ontology. We exploit the ontology to (a) construct new candidate redescriptions (i.e., simulation model structures) that can be realized via the ontology; (b) test the candidate redescriptions in terms of their explanatory value, where "value" can be a plurality of considerations, such as empirical goodness of fit, structural parsimony, interpretability, and theoretical credibility.

The ontology developed by the modeler can be considered as a set of basic building blocks of entities and mechanisms. While we could, as humans, use the building blocks to construct an exhaustive set of possible alternative simulation model structures (by assembling the building blocks in different ways), even a relatively small set of building blocks can result in a very large number of alternatives that cannot be practically explored by hand. An alternative is to use machine learning approaches, where we make intelligent use of computational resources to automatically search through the space of possible model structures. In the vein of the existing literature, we regard the family of evolutionary computing approaches known as *genetic programming* (GP) as a highly promising workhorse for structural calibration [13]. Multiobjective genetic programming is particularly beneficial because it allows researchers to evaluate candidate model structures according to a set of explicitly stated considerations of explanatory value [14]. In our present framework, we concentrate on two aspects of value: (i) the ability of the model structure, with suitably calibrated parameters, to reproduce the phenomenon so far as we understand it from our beliefs and empirical data; (ii) the meaningful interpretability of the model structure in terms of theory.

2.4. Generating Candidate Model Structures: Grammar-Based Genetic Programming.

Evolutionary computing is a field that applies the principles of natural evolution in computing. In EC, a population of candidates is evolved over many generations based on a fitness function. A typical process starts with a random population of candidates. The candidates with high fitness are then probabilistically chosen to breed and produce the candidates for the next generation. Two common genetic operators for breeding are crossover (combining random parts from two selected candidates) and mutation (altering a random part of a selected candidate). Genetic programming applies this idea of evolution for computer programs [13, 15].

The basic genetic operators (i.e., crossover and mutation) are entirely random and can result in the construction of illegal programs (e.g., that breach requirements for legal expressions or type restrictions of the programming language). For this reason alone, it is often appropriate to constrain the structure of programs in advance of the evolutionary process. An approach to enforcing particular structures is using a grammar [16]. GP approaches that use a grammar to express constraints are called grammar-based genetic programming (GGP). For example, the expression $f(x, y) = x * x + y$ is one of many possible specific structures that could be generated with the following grammar:

$$
\begin{aligned}
\mathrm{E} &:= \mathrm{var} \,|\, (\mathrm{E\ op\ E}), \\
\mathrm{op} &:= + \,|\, - \,|\, *, \\
\mathrm{var} &:= \mathrm{x} \,|\, \mathrm{y} \,|\, \mathrm{z}.
\end{aligned}
\tag{1}
$$

Each line in the grammar is a production rule. The elements on the left-hand-side can be rewritten and are called nonterminal symbols. On the other hand, elements that cannot be rewritten are terminals. The first production rule is an expression (E) which can equal either a variable (var) or a combination of two expressions (E) by an operator (op). The second rule allows three operators: plus, minus, and multiply. The last production rule specifies three variables: x, y, and z.

Each structure produced by the grammar is represented by a tree. The tree representation allows researchers to measure the structural complexity of models by counting the number of nodes (terminal and nonterminal symbols) in the tree. Even the simple expression $f(x, y) = x * x + y$, shown in Figure 1, has a node count of 15. Crossover is operationalized by cutting a branch of the tree and replacing it with a branch from another tree. Mutation is operationalized by replacing a node with a randomly generated tree.

2.5. Description of the Model Discovery Process.

This section describes the proposed model discovery process, depicted by the flowchart of Figure 2. The process is a variant of a recent approach described by Vu et al. [10]. There are three roles in
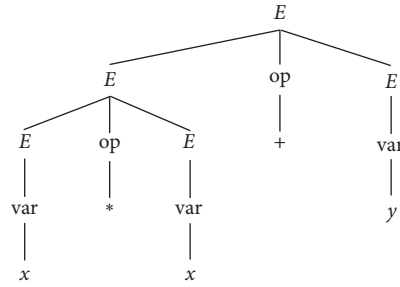
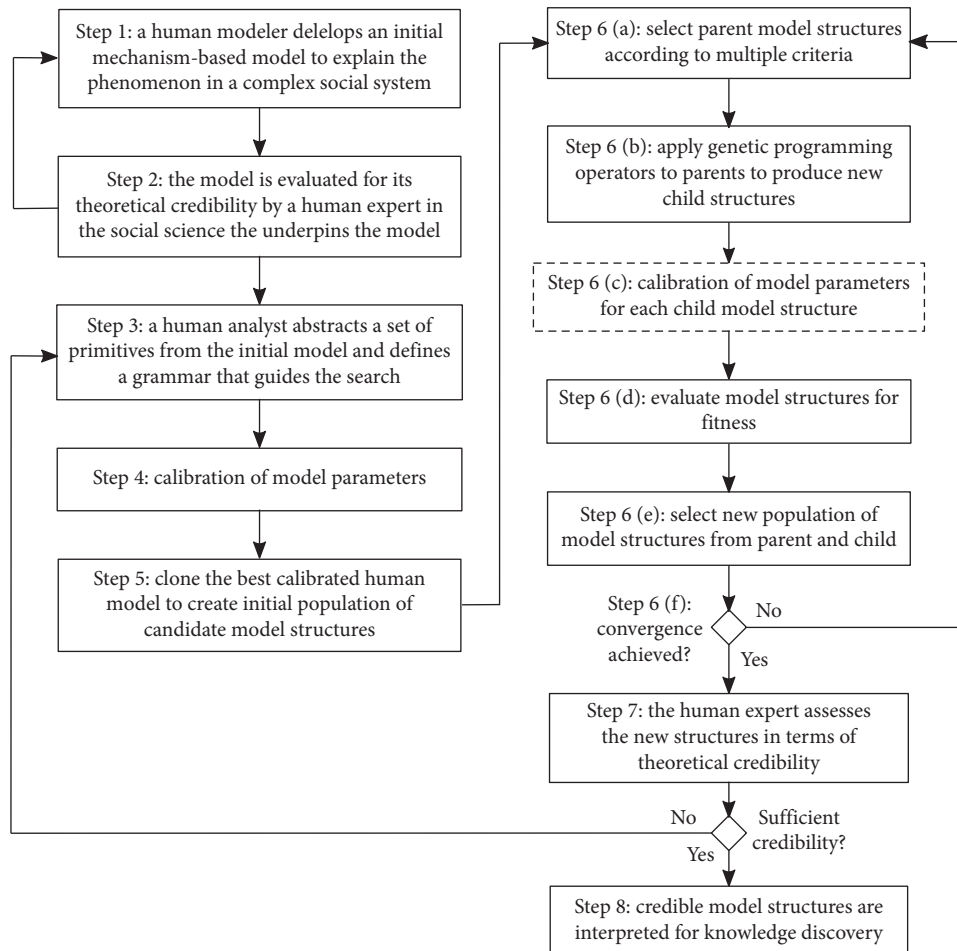FIGURE 1: An example tree of an expression $f(x, y) = x * x + y$.

FIGURE 2: Model discovery framework. The dashed rectangle is the process that is not implemented in the present paper.

the model discovery process: modeler, analyst, and domain expert. The modeler designs, implements, and tests agent-based models. The analyst analyses the model structure and abstracts a set of basic building blocks of entities and mechanisms. The domain expert possesses the knowledge and understanding about the social science that underpins the model and can assess the model's theoretical credibility.

In Step 1 of the model discovery process, a human modeler develops a mechanism-based model to explain the phenomenon in a complex social system (the baseline model). This redescription process is undertaken by the

modeler based on existing knowledge captured in social theories. In Step 2, the model is evaluated for its theoretical credibility by a human expert in the social theories that underpin the model (i.e., a domain expert). Redevelopment of the model may occur following this step (representing a return to Step 1). Once the human expert is satisfied with the baseline model, in Step 3, a human analyst abstracts a set of primitives, i.e., "building blocks," from the model. These sets of primitives are the entities and mechanisms to be exposed and modified in the evolutionary step (Step 6). In addition, a grammar is defined to guide the search. Since the grammar is

a set of production rules for combining the primitives, the analyst can enforce certain structures based on the modeler's knowledge of the system; noncredible operations can be prohibited. In Step 4, the model parameters are calibrated using the baseline model structure against the empirical calibration targets. In Step 5, the model built by human modeler, along with the best calibrated parameters, is cloned to fill the initial population of model structures.

Step 6 is the heart of the evolutionary approach. Parent structures are selected from the population of model structures (Step 6(a)). After applying the genetic operators (crossover and mutation), new child structures are generated (Step 6(b)). Ideally, the parameters of the new child structures are recalibrated to see if the model error can be minimized further (Step 6(c)). However, such a nested approach to calibration is very computationally intensive, and so we necessarily omitted this step due to limits on the available computing resources. Instead, we allowed the GP to select constants from not only a general set of constants but also values of calibrated parameters generated at Step 4. Next (in Step 6(d)), the new structures are evaluated for their fitness (such as model error compared to empirical data). After evaluation, the new population is selected based on the objectives (Step 6(e)). These evolutionary steps are performed until convergence is achieved or when a maximum number of iterations is reached (Step 6(f)).

In Step 7, through deliberative discussion with the analyst and the modeler, the domain expert assesses the set of new structures with the highest fitness values in terms of theoretical credibility. If the new structures lack sufficient credibility, the domain expert, the modeler, and the analyst return to Step 3 to discuss changes to the grammar to improve the meaningfulness of the operations that can be selected by the evolutionary algorithm. Further iterations of Steps 3 to 7 are carried out until credible structures are generated or resources are exhausted. In Step 8, credible model structures are interpreted for knowledge discovery purposes, promoting discussion about the underlying social theories used in the model and, potentially, further theory development or empirical data gathering.

3. Application

We applied the new framework to a specific mechanism-based social systems model. Here, we interpreted the causal mechanisms derived from *social role theory* as drivers of alcohol consumption to build a complex systems model of population-level alcohol use patterns in the US since the 1980s (Figure 3). Social role theory is a collective term used to describe a diverse range of mechanism-based explanations for individual and collective behaviors and practices from the fields of social psychology, sociology, and anthropology [17]. Particular conceptualizations of role theory have been used within the alcohol research community to explain observed trends in alcohol use—specifically relating to the interplay between alcohol use and positional roles such as parent, partner, and paid employee [18]. Our aim in this application is to test the extent to which credible conceptualizations of role theory can reproduce historical

trends in population alcohol use (as measured via survey data).

In this application, different roles in the model discovery process were undertaken by different authors. The modeler role was principally undertaken by HB, AB, and RCP. The analyst role was undertaken by CB and TMV. The domain expert was PS.

3.1. Social Roles as a Mechanism-Based Explanation of Alcohol Use: The Human-Built Model as It Relates to Role Theory. The concept of *role strain* is central to many of the studies relating positional roles to alcohol use. Biddle [17] defines role strain as the "experience of stress associated with positions or expected role." Role strain is hypothesized to arise through a number of pathways where alcohol can act as cause, consequence, or both. Alcohol can be used by individuals as a means of coping with role strain arising from role overload (holding a role set that is too complex), role deprivation (lacking roles that provide meaning to life), or role incongruence (holding roles which are nonnormative with respect to status or identity) [19, 20]. Alcohol use can also induce or exacerbate role strain, where use is incompatible with the demands of performing the role [21]. In the model, role strain is the arithmetic mean of role incongruence and role overload (equation (1) in Table 1). Role overload (equation (2) is determined by the roles an agent holds, their levels of involvement in these roles, and four calibrated parameters representing the effect of holding each role on experiencing role overload. Role incongruence (equation (3)) is the arithmetic mean of the difference between each role holding status and the prevalence of that role in society (i.e., the percentage of people holding that role).

In the mechanism known as *role selection*, individuals may act (consciously or otherwise) to prevent or reduce role strain by avoiding or escaping from roles that are incompatible with their existing alcohol use [21]. This mechanism is implemented by adjusting the probability of transitioning between roles based on the heavy drinking status of the agent (where heavy drinking is defined as having consumed 5+ standard drinks in the previous month) (equations (4) and (5)). In a contrasting *role socialisation* mechanism, individuals gradually adopt and internalize drinking practices that are compatible with the roles they hold [21–23]. A difference in drinking disposition is calculated when individuals gain or lose roles (equations (6) and (7)). The new disposition to drink (equation (8)) is a function of this difference in disposition and a modifier (equation (9)), calculated using the number of days the new role has been held and the speed of socialisation.

Knibbe et al. [18] suggested that the set of positional roles held by an individual can affect the ability of that person to participate in drinking situations, depending on the extent to which drinking is integrated into the structure of everyday life within society. In this sense, social roles act as mechanisms that regulate the daily *opportunities* for using alcohol. Individuals in the model have a different opportunity to drink outside and inside the home, which depends on the roles they hold. The opportunity to drink out and in are each
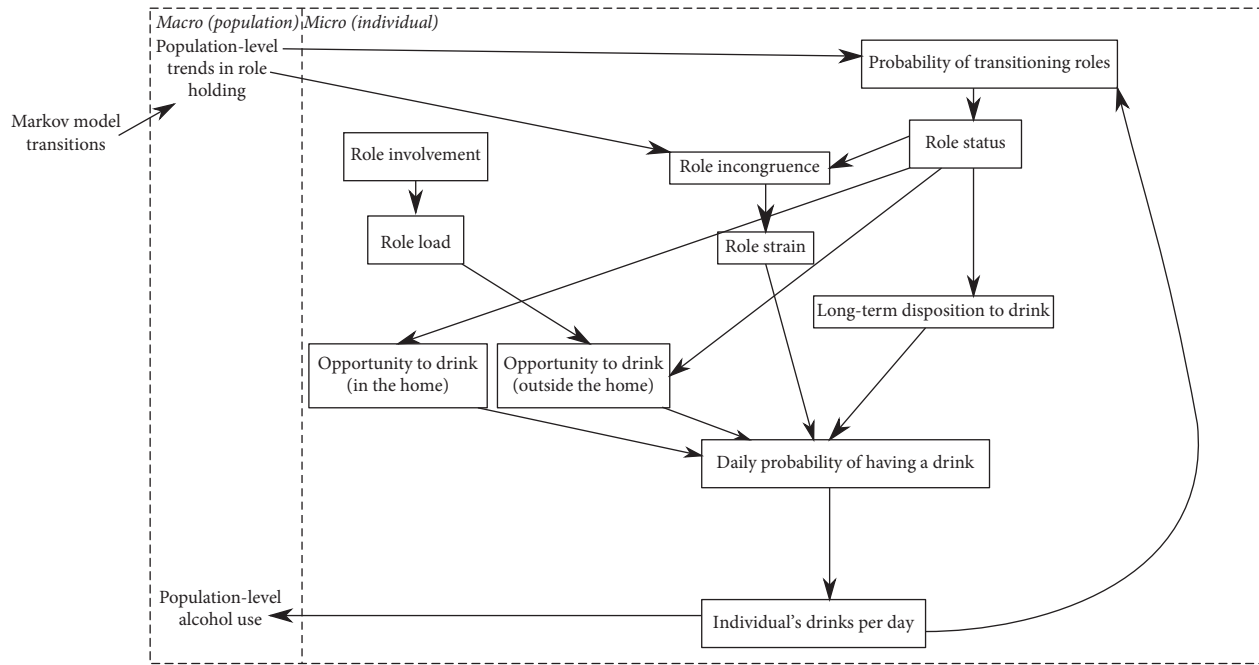
FIGURE 3: Schematic of roles model.

calculated as a log odds (equations (10) and (11)) and then converted to probabilities (equations (12) and (13)). The log odds for drinking outside the home are based on an agent's role load, employment status, and two calibrated parameters to describe the unknown effect sizes of these factors on opportunity to drink out. Role load and a combination of marital and parenthood status determine the log odds of drinking opportunity inside the home. Again, this equation contains parameters describing the unknown effect of these factors on the opportunity to drink in.

The conditional probability of agent i consuming a jth drink (equations (14) and (15)) is governed by each agent's long-term disposition to drink, their probability of drinking in and outside the home, and role strain.

3.2. Data

3.2.1. Model Initialization. To initialize the models, data from the National Survey on Drug Use and Health (NSDUH) 1979–2010 [24], Panel Study of Income Dynamics (PSID) 1979–2010 [25], and the US Census 1980–2010 [26] were used. A microsynthesis [27] was generated for a demographically representative population of 1000 individuals aged 12–80 in the USA, 1980. The model was initialized with these 1000 agents on the first day of 1980. The sociodemographic attributes of agents were initialized from the microsynthesis: age, sex, ethnicity, employment status, marital status, and parenthood status. Additionally, the microsynthesis initialized agents with alcohol use attributes: a 12-month drinking status, usual number of drinking days per month, usual quantity of drinks per month, and number of days where more than five drinks are consumed per month.

3.2.2. Simulation. During each simulated year, individuals enter the model as new 12-year-old adolescents and new migrants. Individuals also leave the model due to death and outward migration. Total counts of new migrants to enter in each year were estimated using the American Community Survey 1980–2010 [28] and were microsynthesised to data from the nearest NSDUH year to give a representative migrant population with corresponding baseline drinking behavior. Mortality rates for the microsimulation were derived from the Center for Disease Control and Prevention (CDC) all cause mortality data for the US between 1979–1998 [29] and 1999–2010 [30].

Transition probabilities for moving between each of the eight unique combinations of social roles variables (marriage, employment, and parenthood) are applied annually during the simulation. These probabilities were derived from multistate Markov models fitted to marriage, parenting, and employment trends from a representative US study, the Panel Study of Income Dynamics 1979–2000 [25].

At initialization, each agent is allocated a vector which represents their long-term disposition to drink. These are initialized from the mean and standard deviation of their drinking frequency and quantity at baseline.

3.2.3. Calibration Targets. Calibration targets for alcohol use were derived from empirical data from NSDUH for the years 1979–2010. Four alcohol use targets per year were used for calibration: (1) prevalence—the proportion of individuals reporting consuming an alcoholic beverage during the previous year; (2) frequency—among drinkers, the average number of drinking days per month; (3) quantity—among drinkers, the average grams of alcohol consumed per day;

TABLE 1: Equations in the human-built model.

No.	Concept	Model equation	Description
1	Role strain	$\text{RoleStrain}_i[k] = (\text{RoleLoad}_i[k] + \text{RoleIncongruence}_i[k])/2$	Role strain is the overall stress an individual experiences as a result of the social roles they hold.
2	Role load	$\text{RoleLoad}_i[k] = \boldsymbol{\beta_1} * \text{ParentStatus}_i[k] * \textbf{ParentInvolvement}_i + \boldsymbol{\beta_2} * \text{MaritalStatus}_i[k] * \textbf{MaritalInvolvement}_i + \boldsymbol{\beta_3} * \text{EmploymentStatus}_i[k] * \textbf{EmploymentInvolvement}_i + \boldsymbol{\beta_4} * (1 - \text{MaritalStatus}_i[k]) * \text{ParentStatus}_i[k] * \textbf{ParentInvolvement}_i$	Role load is the stress that results from needing to perform a role. Role status is either 0 (not having a role) or 1 (having a role). Role involvement represents how much a person is involved in a role, if they hold it (between 0 and 1, from no involvement to full involvement). There are four terms: one term for each of the three roles (having a role and more involvement in that role increases the stress) and a term for additional stress when an individual is a single parent (holding the role without the support of another parent).
3	Role incongruence	$\text{RoleIncongruence}_i[k] = (\text{ParentStatus}_i[k] - \text{sParentExpectancy}_{sex,age}[k] + \text{MaritalStatus}_i[k] - \text{sMaritalExpectancy}_{sex,age}[k] + \text{EmploymentStatus}_i[k] - \text{sEmploymentExpectancy}_{sex,age}[k])/3$	Role incongruence is the stress that results from holding a role that deviates from societal expectations for an individual's identity (encoded as a sex-age category). It is the average of the differences for each role between the current status and the corresponding societal expectancy (prevalence of that role in the society is between 0 and 1).
4	Role transition update for gaining roles	Heavy drinkers: (i) $\text{TP}_i[k] = \text{sTP}_{sex,age}[k-1] * (1 + \boldsymbol{\beta_{12}})$ Non heavy drinkers: (i) $\text{TP}_i[k] = \text{sTP}_{sex,age}[k-1] * (1 - \text{AnnualHeavyDrinkingPrevalence}[k-1] * (1 + \boldsymbol{\beta_{12}}))/(1 - \text{AnnualHeavyDrinkingPrevalence}[k-1])$	To account for role selection, individual role transitions over the lifecourse are calculated by modifying the societal transition rates according to whether or not the individual is a heavy drinker (equations (4) and (5)). Heavy drinking makes it less likely for an individual to gain roles. Heavy drinking makes it more likely for an individual to lose roles. AnnualHeavyDrinkingPrevalence represents the population prevalence of heavy drinking in the model (between 0 and 1).
5	Role transition update for losing roles	Heavy drinkers: (i) $\text{TP}_i[k] = \text{sTP}_{sex,age}[k-1] * (1 + \boldsymbol{\beta_{13}})$ Non heavy drinkers: (i) $\text{TP}_i[k] = \text{sTP}_{sex,age}[k-1] * (1 - \text{AnnualHeavyDrinkingPrevalence}[k-1] * (1 + \boldsymbol{\beta_{13}}))/(1 - \text{AnnualHeavyDrinkingPrevalence}[k-1])$	

TABLE 1: Continued.

No.	Concept	Model equation	Description
6	Difference in disposition to drink due to gaining roles	$\text{DispositionDifference}_{i,j}[k] = \text{Disposition}_{i,j}[k] * (1 + \boldsymbol{\beta_{10}}) - \text{Disposition}_{i,j}[k]$	To account for role socialisation, the disposition to drink is gradually reduced the longer an individual holds a role and is gradually increased if an individual loses a role. The full disposition effect to apply is calculated using equations (6) and (7). The proportion of that effect to apply after a particular number of days of socialisation is calculated using the logistic function in equation (9). This modifier is then applied to scale the full disposition effect using equation (8) and calculate the overall disposition at time k. Socialisation effects accrue over one year following a role transition.
7	Difference in disposition to drink due to losing roles	$\text{DispositionDifference}_{i,j}[k] = \text{Disposition}_{i,j}[k] * (1 + \boldsymbol{\beta_{11}}) - \text{Disposition}_{i,j}[k]$	
8	New disposition to drink (after role socialisation)	$\text{Disposition}_{i,j}[k] = \text{Disposition}_{i,j}[k-1] + \text{DispositionDifference}_{i,j}[k] * \text{modifier}_i[k]$	
9	Modifier for socialisation mechanisms	$\text{modifier}_i[k] = \hat{e}((\text{DaysofSocialisation}_i[k] - \boldsymbol{\text{sSocialisationSpeed}})/365)/ (1 + \hat{e}((\text{DaysofSocialisation}_i[k] - \boldsymbol{\text{sSocialisationSpeed}})/365))$	
10	Opportunity to drink out	$\text{logOddsOppOut}_i[k] = \log(\boldsymbol{\beta_5} * (1 - \boldsymbol{\beta_6} * \text{RoleLoad}_i[k] + \boldsymbol{\beta_7} * \text{EmploymentStatus}_i[k]))$	Equations (10) and (11) describe the log odds for the opportunities to drink outside and inside the home, with reference to having no opportunity to drink. β_5 is the baseline opportunity. Role load acts to reduce both opportunities. Individuals have more opportunity to drink outside the home if employed and more opportunity to drink inside the home when holding marital or parenting roles. Equations (12) and (13) operationalize the logit model that derives the probabilities of drinking outside and inside the home on any given day from the log odds of equations (10) and (11) (for three mutually exclusive scenarios: drinking in, drinking out, and not drinking)
11	Opportunity to drink in	$\text{logOddsOppIn}_i[k] = \log(\boldsymbol{\beta_5} * (1 - \boldsymbol{\beta_8} * \text{RoleLoad}_i[k] + \boldsymbol{\beta_9} * (\text{MaritalStatus}_i[k] + \text{ParentStatus}_i[k])))$	
12	Probability of having an opportunity to drink out	$\text{probOppOut}_i[k] = \hat{e}(\text{logOddsOppOut}_i[k])/ ((\hat{e}(\text{logOddsOppIn}_i[k]) + \hat{e}(\text{logOddsOppOut}_i[k]) + 1))$	
13	Probability of having an opportunity to drink in	$\text{probOppIn}_i[k] = \hat{e}(\text{logOddsOppIn}_i[k])/ ((\hat{e}(\text{logOddsOppIn}_i[k]) + \hat{e}(\text{logOddsOppOut}_i[k]) + 1))$	
14	Probability of drinking first drink ($j = 0$)	$\text{ProbabilityDrink}_{i,0}[k] = \text{Disposition}_i[k] * (\text{ProbOppOut}_i[k] + \text{ProbOppIn}_i[k]) * (1 + \boldsymbol{\beta_{14}} * \text{RoleStrain}_i[k])$	The daily drinking probability is modeled as the long-term drinking disposition, mediated by drinking opportunities and role strain. We differentiate between drinking frequency (first drink in an occasion) and quantity (next drink, given that an occasion has begun).
15	Probability of drinking next drink ($j > 0$)	$\text{ProbabilityDrink}_{i,j}[k] = \text{Disposition}_i[k] * (\text{ProbOppOut}_i[k] + \text{ProbOppIn}_i[k]) * (1 + \boldsymbol{\beta_{15}} * \text{RoleStrain}_i[k])$	

The concepts in the concept column are from the schematic in Figure 3. These equations contain unobserved parameters (highlighted in bold) which modify the effects of social role mechanisms. These are given values following model calibration (Section 3.3) which searches for the set of parameters which best fit historically observed trends in alcohol use over time. The agents in the model, indexed by i, represent individual drinkers; their behavior is a decision to consume the jth drink in a drinking occasion. The model also includes two dynamic structural entities: expectancies for holding roles at a given point in the lifecourse and average transition probabilities (TPs) between roles. For clarity, all structural entities carry the prefix s. The discrete time unit (representing each day) in the simulation is k.

and (4) heavy episodic drinking—among drinkers, the average number of occasions where 5+ drinks were consumed, per month. The targets are split by subgroup, with four subgroups defined by the number of roles held (0–3 roles). We chose to categorize by the number of roles ($n = 4$) instead of the combination of roles ($n = 8$) for two reasons: firstly, this

is an indicator commonly used in the social roles literature [19]; secondly, the eight role decomposition is too great for the standard error of the targets to be informative from a calibration perspective. In summary, there are 16 targets (4 alcohol use targets by 4 different number of roles) for each year between 1979 and 2010.

3.2.4. Implementation. The model was implemented in C++ using the RepastHPC 2.2.0 toolkit [31]. The model is run forward in time for 20 years for calibration (1980–1999) and 10 years for validation (2000–2009). Each model tick represents one day in the simulation. On each day of the simulation, the probability of drinking is calculated for each agent. Once per year in the simulation, transition probabilities for role transitions are applied and role expectancies are updated.

3.3. Parameter Calibration. The model contains 31 parameters for calibration, which are highlighted in bold text in Table 1. For this paper, a Latin hypercube space-filling design was employed to sample 5,000 parameter sets from the joint prior distribution using the lhs package in *R* [32]. The Latin hypercube was optimized by maximizing the minimum distance between samples [33]. These parameter sets are evaluated using an error metric that compares the simulated results against the calibration targets.

$$\text{error} = \frac{1}{\text{NM}} \sum_{n=1}^{N} \sum_{m=1}^{M} \frac{|y_m^*[n] - y_m[n]|}{\sqrt{(s_m[n])^2 + (d_m)^2}}, \quad (2)$$

where N is the number of observations, M is the number of outputs, $y_m^*[n]$ is the simulated data for output m at time n, $y_m[n]$ is the mean of empirical target data for output m at time n, $s_m[n]$ is the standard error of the empirical target data for output m at time n, and $(d_m)^2$ is the variance of the model discrepancy for output m, which is 10% of the possible range for each output. Model discrepancy captures the fact that model is not a perfect representation of reality.

As described in Section 3.2.3, there are 16 targets (prevalence, frequency, quantity, and heavy episodic drinking, each split by the number of roles held, 0–3), and thus M is 16. N is different for each output because some years are missing in the empirical data. The parameterization that provided the minimum error in equation (2) was selected as the result of the calibration process. The human-built model along with the best parameterization was selected as the reference model for the structural calibration process.

3.4. Grammar-Based Genetic Programming: Design and Implementation. This section describes a GGP system designed to perform the model discovery process. For the grammar that guided the GP process, the popular context-free grammar was used. The full grammar written in Backus–Naur form [34] is available in Figure 4. Considering the representation of the GGP candidates, each candidate is represented by a tree that is generated following the production rules of the grammar. Each GGP candidate (a program <p>) contains 9 expressions for the 9 role-related terms used in the roles model: role selection mechanism, role socialisation mechanism, role load, role incongruence, role strain, log-odds-out modifier, log-odds-in modifier, first-drink disposition, and next-drink disposition. Several groups of variables, along with constants and calibrated parameters, were defined. Each can be formed only by a defined combination of variables, operators, and constants.

Some role-related terms use the same expression because they have the same structure and the same constraints, e.g., `probFirstDrink` and `probNextDrink` both use expression <e5>. This grammar provides a hierarchical structure that can capture the layers of role-related concepts in the reference model.

For the initialization of the GGP, we decided to start with an initial population filled with the same structure. The structure used as the starting point is the reference model, i.e., the structure designed by human modelers with the best fitted parameterization. Additionally, a multiobjective GGP was employed to simultaneously minimize both model error and complexity. These two objectives address the two aspects of value we discussed in Section 2.3: (i) the ability of the model to reproduce the phenomenon so far as we understand it from our beliefs and empirical data; (ii) the meaningful interpretability in terms of theory.

The first objective, model error, is captured by comparing the simulated data from the model with the empirical data from the real world. The model error is described in equation (2) (Section 3.3). The second objective, complexity, is a proxy for interpretability and parsimony. Minimizing the complexity during model discovery also constrains the model discovery process from discovering too complex structures that overfit the empirical data and are not interpretable by domain experts. The complexity is defined by the number of nodes in the GGP candidate, with a special case that node ON is counted as 2 in contrast to node OFF being counted as 1. The drawback when using complexity as a proxy is that it does not guarantee meaningful interpretability, i.e., low complexity can also result in meaningless model structures. Therefore, at the end of every iteration of the model discovery process, we worked with the domain expert to verify the interpretability of all structures in terms of theoretical meaningfulness. We asked the expert to classify model structures using a crisp binary definition of credible or noncredible. While the judgment was holistic, the classification process was a deliberative discussion between the expert, the modeler, and the analyst. This discussion was recorded and used subsequently to produce a set of qualitative criteria for judging model credibility.

For the selection process, the popular NSGA-II optimizer [35] was used to develop an even representation of the Pareto front that shows the trade-off between model error and complexity. During the evaluation, the corresponding expressions in the simulation's source code were edited based on a candidate structure; then the simulation was recompiled and run in order to collect the simulated results for calculating the model error.

The described GGP system was implemented using the PonyGE2 toolkit [34] and set up with the following parameters:

(i) 500 GGP candidates per generation

(ii) Initialization: 500 copies of the reference structure

(iii) GP operators: 75% subtree crossover and 25% subtree mutation

(iv) Maximum tree depth: 17

```
<p> ::= RoleSelectionMech=<e0>; RoleSocialisationMech=<e0>; RoleLoad=<e1>;
RoleIncongruence=<e2>; RoleStrain=<e3>; OutMod=<e4>; InMod=<e4>;
probFirstDrink=<e5>; probNextDrink=<e5>;



# Variables
<v0> ::= MaritalStatus | ParentStatus | EmploymentStatus
<v1> ::= InvolvedxMarital | InvolvedxParent | InvolvedxEmployment
<v2> ::= DiffExpectancyMarital | DiffExpectancyParent |
DiffExpectancyEmployment
<v3> ::= RoleLoad | RoleIncongruence
<v4> ::= RoleStrain | ProbOppIn | ProbOppOut | Disposition



# Toggle for mechanisms
<e0> ::= ON | OFF

# Role load
<e1> ::= (<e1>+<e1>) | (<e1>-<e1>) | (<e1>*<e1>) | (<c>*<e1>) | (<c>+<e1>) |
(1-<e1>) | (<e1>-<c>) | <v1> | <v0>

# Role incongruence
<e2> ::= (<e2>+<e2>) | (<e2>-<e2>) | (<e2>*<e2>) | (<c>*<e2>) | (<c>+<e2>) |
(1-<e2>) | (<e2>-<c>) | <v2> | (<v0>*<e2>) | ((1-<v0>)*<e2>)

# Role strain
<e3> ::= (<e3>+<e3>) | (<e3>-<e3>) | (<e3>*<e3>) | (<c>*<e3>) | (<c>+<e3>) |
(1-<e3>) | (<e3>-<c>) | <v3>

# Modifiers for logOddsOppOut/In
<e4> ::= (<e4>+<e4>) | (<e4>-<e4>) | (<e4>*<e4>) | (<c>*<e4>) | (<c>+<e4>) |
(1-<e4>) | (<e4>-<c>) | <v0> | RoleLoad

# Daily drinking probability (first and next drink)
<e5> ::= (<e5>+<e5>) | (<e5>-<e5>) | (<e5>*<e5>) | (<c>*<e5>) | (<c>+<e5>) |
(1-<e5>) | (<e5>-<c>) | <v4>



# Constants and calibrated parameters
<c> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
        0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 |
        10 | 100 | 0.01 | 0.001 | 0.333 | 0.25 |
        0.179 | 0.038 | 0.128 | 60.952 |
        0.253 | 11.187 | 23.786 | 2.010 |
        1.104 | 9.100
```

Figure 4: Context-free grammar. Variables have the same name as in Table 1 but, for improved clarity, without the indexing notation. Shorthand notation is defined for the multiplication of role involvement and role status (e.g., InvolvedxMarital = MaritalInvolvement * MaritalStatus) and for the difference between the role status and role expectancy (e.g., DiffExpectancyMarital = MaritalStatus − sMaritalExpectancy).

(v) 2 objectives (goodness of fit and complexity) with NSGA-II replacement and selection operators

The GGP process was run on an Intel i9 9980XE processor with 36 cores. The source code of the simulation with the best calibrated parameters (RepastHPC) and the GGP system (PonyGE2) is available at bitbucket.org/r01cascade/ roles_ggp_complexity and is licensed under the GNU General Public License version 3.

4. Results

4.1. GGP Results and the Pareto Front. In the case study, three iterations of the model discovery process were

required to produce any structures that the domain expert deemed as credible. Modifications were made to the grammar between each iteration in an attempt to improve the effectiveness of the discovery process. This was an open-ended trial and error iterative process involving the modeler and the analyst. We stopped the process once credible structures had been discovered. The evolution of the grammar is documented in the Supplementary Material A. The final grammar is shown in Figure 4.

In the final iteration of the process, both model error and model complexity objectives reduced over generations and converged at the 20th generation, after which no change to the Pareto front was observed. Figure 5 shows the final population of 14 nondominated structures and also includes the reference structure for comparison. These models are indexed by their complexity, e.g., model 24 is the model on the Pareto front with complexity 24. All the structures discovered by the GGP are less complex than the reference structure. Six of them are worse than the reference model in terms of model error, while the remaining eight offer improved fit over the calibration window.

4.2. Theoretical Credibility of the Discovered Models. We worked with a domain specialist to examine the non-dominated model structures generated by the GGP in terms of their theoretical credibility and coherency with respect to social role theory. Table 2 compares the structures of the reference model and selected GGP models. In Table 2, elements not affecting agent drinking (`probFirstDrink` and `probNextDrink` variables, equations (14) and (15)) are highlighted in italic.

Through analysis of the deliberative discussions held between the expert, analyst, and modeler, it was possible to generate three qualitative criteria that enable a consistent assessment of credibility for this example case study. In what follows, we provide a narrative discussion of credibility across the Pareto front in Figure 5. However, a complete documentation of all GGP models and corresponding justifications of credibility is also included in the Supplementary Material B. The three identified criteria for theoretical credibility are as follows:

(1) At least one of the theory constructs must be implicated in the model dynamics. In a mechanism-based model of alcohol use, the mechanisms need to be used to generate drinking behavior. For models based on role theory, this means that the models must use at least one of the core theory constructs of role strain, role load, role incongruence, or opportunity. For example, in the absence of a role socialisation process, using solely a variable that describes drinking history (`Disposition`) does not generate a credible mechanism-based model in terms of role theory.

(2) The theory constructs must be used to represent mechanisms, rather than being proxies for black-box variable-centric explanation. In some of the
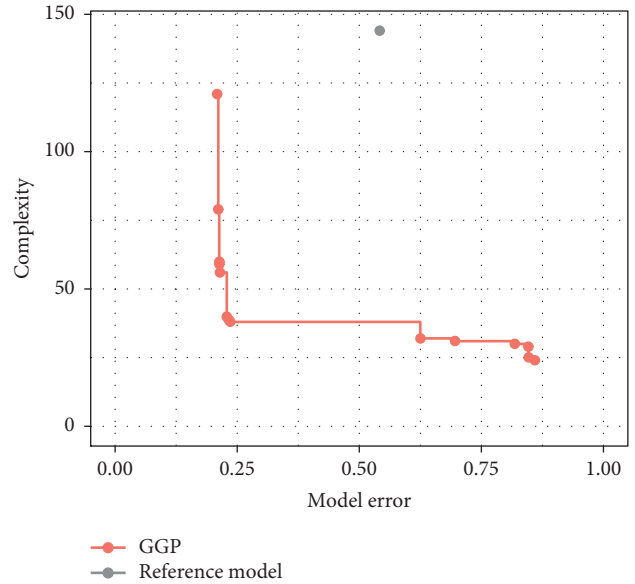

Figure 5: GGP structures on the Pareto front versus the reference structure (model error is the error when compared the simulated results against calibration targets; complexity is measured as the number of nodes in the tree of a GGP candidate).

identified models, we observed that theory constructs could be replaced directly by observable sociodemographic properties of the agents. These cases indicate that the mechanism-based explanation is being avoided in favor of a black-box variable-centric "determinants" approach to understanding drinking behaviors that is more conventional in the literature [36]. For example, in a mechanism-based model, marital status should not directly define opportunity, where opportunity then directly determines disposition to drink.

(3) The model equations that describe the mechanisms must be compatible with the causal logic and evidence base for the theory. In a mechanism-based model, some of the encoded causal relationships between core theory constructs are only meaningful when constrained in terms of direction, sign, and/or magnitude. For example, role load must either not affect or cause a decrease in opportunity to drink—it is inconceivable, in role theory terms, that role load could cause an increase in opportunity (since load implies time use by agents that cannot be combined with drinking).

Focusing now on the Pareto front, the model with lowest complexity on the front (on the far right of Figure 5) is model 24—this model includes a single term for each production rule, with both role selection and role socialisation processes switched off (Table 3). When parsed, all but two of the production rules are inactive—`probFirstDrink` and `probNextDrink`—with both set to the term `Disposition`. This term represents the initial dispositions to drink endowed to the agents based on observed drinking patterns in NSDUH data, so essentially model 24 encodes the

TABLE 2: Structures of the discussed models (structures of the reference model, model 24, model 25, and model 29).

Name	The reference model	Model 24	Model 25	Model 29
RoleSelectionMech	ON	OFF	ON	OFF
RoleSocialisationMech	ON	OFF	OFF	OFF
RoleLoad	$0.180 * \text{InvolvedxMarital} + 0.039 * \text{InvolvedxParent} + 0.128 *$ InvolvedxEmployment $+ 60.952 * (1 - \text{MaritalStatus}) * \text{InvolvedxParent}$	*InvolvedxEmployment*	*InvolvedxEmployment*	InvolvedxMarital
RoleIncongruence	$0.333 * \text{DiffExpectancyMarital} + 0.333 * \text{DiffExpectancyParent} + 0.333 * \text{DiffExpectancyEmployment}$	*DiffExpectancyMarital*	*DiffExpectancyMarital*	DiffExpectancyMarital
RoleStrain	$0.5 * \text{RoleLoad} + 0.5 * \text{RoleIncongruence}$	*RoleLoad*	*RoleLoad*	RoleIncongruence
OutMod	$\text{OutMod} = (1 - 0.25357 * \text{RoleLoad}) + 11.18756 * \text{EmploymentStatus}$	*RoleLoad*	*RoleLoad*	EmploymentStatus
InMod	$\text{InMod} = (1 - 23.786 * \text{RoleLoad}) + 2.010 * (\text{MaritalStatus} + \text{ParentStatus})$	*RoleLoad*	*RoleLoad*	RoleLoad
probFirstDrink	$(\text{ProbOppIn} + \text{ProbOppOut}) * \text{disposition} * (1 + 1.104 * \text{RoleStrain})$	Disposition	Disposition	Disposition $*$ ProbOppIn
probNextDrink	$(\text{ProbOppIn} + \text{ProbOppOut}) * \text{Disposition} * (1 + 9.100 * \text{RoleStrain})$	Disposition	Disposition	Disposition

Elements not affecting agent drinking are highlighted in italic.

TABLE 3: Structures of the discussed models (structures of model 38, model 59, and model 79).

Name	Model 38	Model 59	Model 79
RoleSelectionMech	ON	ON	OFF
RoleSocialisationMech	OFF	ON	OFF
RoleLoad	InvolvedxMarital	*InvolvedxMarital*	InvolvedxEmployment
RoleIncongruence	*1 − DiffExpectancyEmployment*	*0.333 ∗ DiffExpectancyMarital + 0.333 ∗ DiffExpectancyParent + 0.333 ∗ DiffExpectancyEmployment*	DiffExpectancyEmployment
RoleStrain	*RoleIncongruence*	*RoleIncongruence*	$7 ∗ RoleIncongruence + RoleLoad$
OutMod	RoleLoad	MaritalStatus	$1 − 10 ∗ RoleLoad$
InMod	EmploymentStatus	ParentStatus	$1 − ParentStatus ∗ RoleLoad$
probFirstDrink	$(LifetimeDisposition^2) ∗ (4 + ProbOppIn)$	$(LifetimeDisposition^2) ∗ (4 + ProbOppIn)$	$Disposition ∗ (0.9 ∗ Disposition + ProbOppIn) ∗ (1.104 ∗ RoleStrain + 0.8)$
probNextDrink	Disposition	Disposition	$1.25 ∗ Disposition2 ∗ (RoleStrain + 1)$

Elements not affecting agent drinking are highlighted in italic.

heuristic "past behavior predicts future behavior" with no aspect of role theory present. This heuristic is clearly sufficient to reproduce target data at the start of the simulation but the fit to targets becomes progressively poor over time.

As we begin to traverse the Pareto front in the direction of increasing complexity (from right to left in Figure 5), elements relating to role theory begin to be introduced into the production rules; however, these elements do not necessarily survive the parsing process. For example, model 25 switches on role selection, but—as with model 24—no components relating to roles are present in the final `probFirstDrink` and `probNextDrink` production rules. We can conclude that the very small improvement observed for model 25 in comparison to model 24 is due to low level stochasticity in the simulation.

In the next model, model 29, role selection and socialisation processes are switched off but production rule `probFirstDrink` is now set to `Disposition * ProbOppIn`, i.e., the probability of engaging in a drinking occasion is scaled by the probability of having the opportunity to drink at home, while the number of drinks consumed in such an occasion continues to follow the heuristic "past behavior predicts future behavior." Following the production rules upward, we identify that, as a result of the log-odds structure that is preserved in all models, `ProbOppIn` is increased by role load (`RoleLoad`) and decreased by holding an employment role (`EmploymentStatus`). Meanwhile, role load is defined as level of involvement in a held marital role (`InvolvedxMarital`). From the perspective of role theory, this model is interpretable but not credible: (i) since only `ProbOppIn` is included, but the possibility of `ProbOppOut = 0` across all agents is not credible, `ProbOppIn` is interpreted as simply a surrogate for any kind of drinking opportunity; (ii) opportunity is seen to increase as a result of role load, which is not credible, since opportunity should decrease with role load, and is seen to decrease as a result of being employed, which is also not credible because, outside of role load considerations (for which employment is not present in the model), being employed should provide increased opportunities for drinking (e.g., due to income or exposure to social drinking situations). Model 29 also offers little improvement in model error compared to the overall "past behavior predicts future behavior." Overall, it is very clear that this model is found wanting.

Continuing to traverse the Pareto front from right to left, we find that the first model to offer a credible interpretation in terms of role theory is model 38. This model also offers a substantial improvement over the less complex "past behavior predicts future behavior" model in terms of model error. In model 38, the frequency of drinking occasions (`probFirstDrink`) is increased by the probability of an at-home drinking opportunity (where, as for model 29, this should be interpreted as a surrogate for any kind of drinking opportunity). Some attempts by the GGP at parameter calibration are also seen here, with nonlinear scaling of `Disposition`. Opportunity is increased by holding an employment role and decreased by role load (the latter defined as level of involvement in a marital role). Despite reservations about the limited extent of the definition of role

load, this opportunity mechanism appears plausible. Role holding is also influenced by drinking behavior in this model, since role selection is switched on—feedback is therefore in action and we can claim at this point that the GGP has discovered a true dynamical model that involves roles.

The second model that can be regarded as credible in terms of role theory is model 59—this model suggests that experiencing role strain increases the likelihood of a drinking occasion (due to the `RoleStrain` multiplier on `Disposition` in the `probFirstDrink` production rule). Role strain is defined purely as role incongruence, where the latter concept is preserved intact from the reference model (i.e., role incongruence is the average deviation of an agent's role holding from normative roles). Both role selection and socialisation are also active in this model, but no opportunity mechanisms are present.

The third model offering credibility in terms of role theory is model 79. In the model, role strain increases both frequency of drinking occasions and per-occasion quantity (via positive modifiers on `probFirstDrink` and `probNextDrink`, respectively). Role strain arises as a weighted combination of role load and role incongruence, where load arises from high levels of involvement in an employment role and incongruence arises through nonnormative employment status (e.g., being working age unemployed). Opportunity also influences drinking frequency (via the `PropOppIn` shift in the `probFirstDrink` production rule)—again, to be credible, `PropOppIn` must be interpreted as a general opportunity, rather than having any locational context. Opportunity is reduced through the interaction of role load with parenting (via the `InMod` production rule) and also reduced by role load in isolation (via `OutMod`). In this model, neither role selection nor socialisation is active.

4.3. Calibrated Goodness of Fit. The reference model has an error of 0.54 against the time series of targets used for calibration (covering the period 1980–2000). The time series plot for the reference model is shown by the pink line in Figure 6. Fit to drinking prevalence is good, except for the one-role subgroup. Fit to frequency, quantity, and heavy episodic drinking is generally poor, particularly for the zero-role subgroup. Models lying on the Pareto front represent a range of errors, including both better and worse than the reference model. The credible models are quite considerably better, as seen from the time series plots in Figure 6. The issue with drinking prevalence in the one-role subgroup and issues with frequency and quantity are largely eliminated—with remaining problems largely confined to underestimation of heavy episodic drinking in the three-roles group.

It should be noted that the goodness of fit of the reference model is weak in comparison to other models. This issue could have been caused by either an inadequate parameter calibration process or fundamental inability of the structures initially designed by the modeler to capture the target dynamics. To seek an improvement in the goodness of fit of the reference model, we could have run a more extensive parameter calibration or undertaken handcrafting of
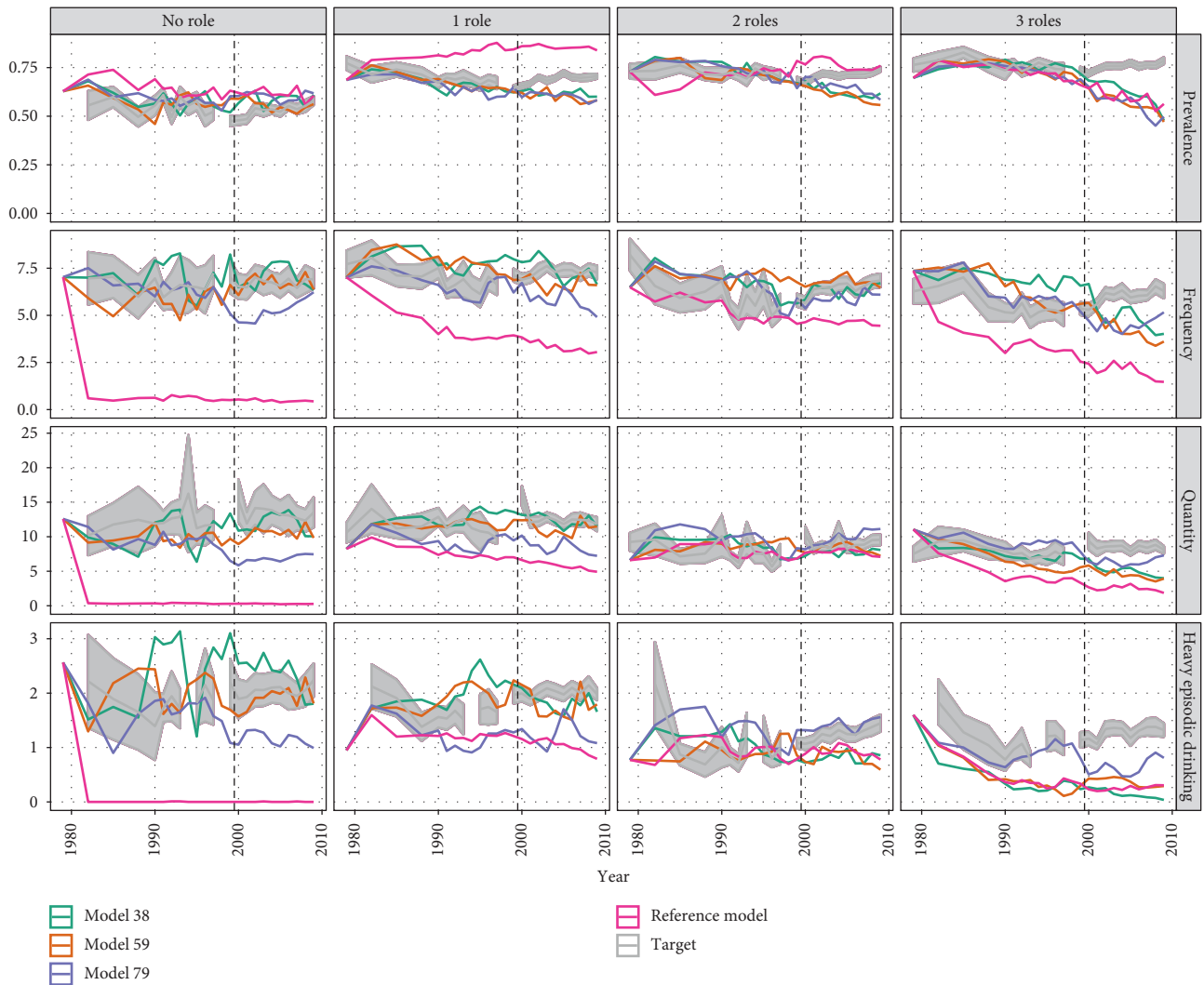
FIGURE 6: Time series of 16 outputs of the reference model, 3 credible GGP models, and the empirical target data (mean target data ± 95% confidence interval). The horizontal axis is the simulated *years* and each graph has a dashed vertical line that separates 20 calibration years and 10 validation years. The vertical axis is the value for each target with different units for *prevalence* (of current drinking in the past 12 months), *frequency* (average number of drinking days in the past 30 days among current drinkers), *quantity* (average grams of pure alcohol consumed per day among current drinkers in the past 30 days), and *heavy episodic drinking* (average number of days where 5+ drinks were consumed in the past 30 days among current drinkers).

the structure. However, we decided the calibrated model was adequate for the GGP process to work with and intentionally did not try to improve the goodness of fit further. It is clear here, from a model development lifecycle perspective, that the boundary between the reference model and the GGP can be blurred.

4.4. Validated Goodness of Fit. Target time series data used for validation covers the period 2000–2010. This period covers an increase in drinking prevalence, frequency, and heavy episodic consumption that contrasts with the gentle declines seen over the calibration window. All three theoretically credible models exhibit a substantial relative decline in performance over the validation period (see Table 4). The calibration and validation errors are inversely correlated,

suggestive of overfitting to noise in the calibration targets. However, the decline in even the lowest complexity credible model (model 38) suggests that the retroduction fundamentally lacks generalizability to an adjacent temporal period, i.e., the real entities and mechanisms identified are invalid or incomplete. Looking in more detail at the validation issues, models 38 and 59 generate continuing declines in drinking for two-role and three-role groups that are trending in the opposite direction in the empirical data. The models also generate a collapse in heavy episodic drinking among the three-role subgroup that is not supported by the data. The most complex of the credible models—model 79—does capture the trend reversal (if slightly lagged) for most two-role and three-role outcomes but underestimates frequency and quantity for the no-role group.

TABLE 4: Goodness-of-fit errors for the reference model and the three credible GP models.

Model	Calibration error	Validation error
The reference model	0.542	0.750
Model 38	0.236	0.340
Model 59	0.214	0.390
Model 79	0.211	0.444

5. Discussion

5.1. Model Discovery Case Study Findings

5.1.1. Insights into Mechanisms. Our model discovery framework offers three alternative perspectives (corresponding to three credible models 38, 59, and 79) that all offer a substantially better fit to the calibration window data compared to the reference model. In perspective (i), the retroducted mechanisms influencing alcohol use are opportunity and role selection; roles affect drinking frequency only, but not quantity. The relevant roles are employment (which drives opportunity) and being strongly involved in a marital role (which reduces opportunity via the role load it creates). Role strain is not important nor is role socialisation. Perspective (ii) suggests that role strain drives increased drinking frequency but not quantity. Role strain is due to holding nonnormative roles, with all three roles implicated. Both role selection and socialisation are important, but drinking opportunity is not important. Finally in perspective (iii), role strain drives both increased frequency and increased quantity. Opportunity drives frequency only. Role selection and socialisation are not important. Parenting and employment are implicated for role strain, but marriage is not. A universal caveat on these perspectives, which were driven by data over 1980–2000, is that validation issues were identified over the period 2000–2010.

5.1.2. How Do These Retroducted Insights Compare to Empirical Findings? In the first perspective (i), holding an employment role increases opportunity to drink, while high levels of involvement in a marital role reduces opportunity to drink, with no other role-based pathways activated. These retroducted mechanisms are both supported by empirical research. Using data from a large birth cohort in the UK, Staff et al. [37] demonstrated that the employment role in isolation was associated with increased alcohol consumption, and both marital and parental roles were associated with a decrease in consumption. Further, the authors suggested that these effects may be caused by differences in opportunities to drink associated with the marital role, for example, by reducing the number of occasions an individual will engage in socialising which could impact alcohol use. This potential mechanism is reflected in perspective (i), whereby an individual has less opportunity to drink and therefore less frequent drinking occasions if they are married (and highly involved in their marital role). This is also supported by Bachman's analysis [22], which found that the association between the marital role and reductions in alcohol consumption was strongly moderated by reductions in evenings out and increased disapproval for use.

In perspective (ii), role incongruence is suggested to be a driver of role strain, which influences the frequency of drinking. This is supported by Biddle [17] who suggests that individuals can experience role strain due to experiencing roles outside the normative timings, for example, transitioning into a parent role as an unmarried teenager. Role strain as a driver of alcohol use has also been empirically observed [19], which suggests that individuals may use alcohol to cope with stress. Additionally in perspective (ii), both role socialisation and selection mechanisms are active. The importance of role socialisation mechanisms as a driver of alcohol use is supported by Lee et al. [23] who found that heavy drinking occasions were reduced after individuals had become married. Additionally, Bachman [22] conducted a review of the literature linking marriage and alcohol use and suggested that the majority of studies find socialisation effects for the marriage role, i.e., gaining the marriage role leads to reductions in alcohol use. The involvement of role selection mechanisms in alcohol use behavior is also supported by the wider literature on role theory and alcohol use. Specifically, Lee et al. [38] provided evidence for role selection mechanisms, finding that earlier alcohol misuse reduces the likelihood of transitioning into social roles. This could be interpreted in a role selection context—if an individual is a heavy drinker, they are less likely to transition into a role which would be incompatible with their drinking.

An increase in alcohol consumption due to role strain is also implicated as a mechanism in perspective (iii). Here, role strain arises due to a combination of role load and role incongruence, which are determined by high levels of involvement in an employment role or a nonnormative employment status, respectively. Role strain as a driver of both alcohol consumption frequency and quantity is supported by Cooper [39], who found in a study of adolescents in the US that drinking as a means of coping (with role strain or otherwise) was associated with heavier drinking patterns. Additionally, in this model, having an opportunity to drink affects frequency of alcohol consumption and is reduced if the parenting role is held. This is supported by Kuntsche and colleagues [19] who found that in a large study of westernised countries, holding a greater number of roles, including parenthood, was associated with a reduction in alcohol consumption, via a decrease in opportunities to drink.

5.2. Benefits of the Framework.

Our approach represents a novel and promising technique for knowledge discovery which is able to generate models with theoretically interpretable mechanisms and can fit historical patterns of data in complex dynamic representations of social systems. The model discovery framework offers a substantial improvement compared to the reference model, in terms of both lowered complexity for interpretability and improved fit to historically observed alcohol consumption trends. However, it is important that in the last step, domain experts are involved to interpret the options generated by the model.

Our technique can therefore provide new approaches for developing theories to explain complex social systems.

A further advantage of this approach is that it is both theory and data-driven. We use formalised theories of behavior and several large empirical datasets to inform both the initial settings of the model (agent population characteristics) and the phenomenon to be explained—population-level alcohol use is derived from a large nationally representative survey.

This method is also very flexible. Firstly, the grammar can be easily modified to redefine search directions, introduce new building blocks, restrict or relax constraints, or introduce different ways to combine the building blocks. This can be done within the grammar without changing the whole model discovery process. Additionally, although we present a case study modeling alcohol consumption, the framework could also be utilized to explain and understand a variety of complex models of social systems. Our method is easily adapted to look at alternative theories of behavior and even to search across multiple theories to give novel combinations which provide a better explanation of empirical data trends.

5.3. Limitations. One limitation of this approach is that the GGP works with the primitives and the grammar that the modeler provides it with. It is therefore possible that theoretically meaningful and adequately explanatory models could be missed because the modeler does not allow for it. In our case study, deliberative discussions during GGP identified that the concept of opportunity encapsulated both time and money resources, which are impacted differently by role theory mechanisms; including these aspects explicitly might arguably have improved the model's explanatory capability. Model discovery is an iterative, problem-specific process. To design the primitives, modelers have to decide which elements in their models are interesting and relevant to their research questions. The level of abstraction is also important: lower levels of abstraction usually have more elements and possible combinations. As for the grammar design, good practices can be found in the work of Nicolau and Agapitos [40].

Additionally, not all aspects of the model were exposed to the GGP process; for example, socialisation and selection mechanisms could be either switched on or off, but the equations could not be modified. If socialisation is switched on, the new disposition to drink is always determined by the same calculation; however, exposing this to the GGP could offer alternative candidate mechanisms for the effects of transitioning roles on underlying desire to drink. This would also allow us to investigate in future model iterations whether socialisation effects vary for different roles, as suggested by Bachman et al. [22].

The GGP method can produce complicated models in terms of theories, which require interpreting by a domain expert. In this paper, out of 14 nondominated structures discovered in the final iteration of the GGP, three were deemed to be theoretically meaningful. For a more efficient search, it would be beneficial if either the grammar allows

only meaningful structures or the model discovery process can enforce the theoretical meaningfulness in other ways (such as during the crossover and mutation operators). However, the prior encoding of meaning is very challenging to achieve and there is also a risk of missing novel ideas due to overconstraining the search. Further, we stopped the discovery process once a small number of credible models had been identified. If we had continued to refine the grammar, further credible models may have been identified—including models that offered reduced complexity or increased goodness of fit over those existing models. In the context of retroduction, such models may have offered greater insight into the relationship between social roles and alcohol use. At present, it remains unclear what a good yield of interpretable theories would be from a model discovery process.

We identified three qualitative criteria for model credibility. These criteria arose from deliberative discussions in relation to the role theory case study and, as such, their generalizability remains untested. We worked with only a single domain expert, but multiple experts may have offered different criteria or interpreted the same criteria differently. While we enforced a crisp binary categorization of model credibility in the search process, given the qualitative nature of the criteria, it may be more appropriate to adopt multinomial and/or fuzzy measures of credibility in future work.

Lastly, our implementation of the proposed discovery process skipped the recalibration stage for parameters of the newly discovered structures (Step 6(c)) due to computational limitations. We addressed this issue by allowing calibrated constants as primitives, but this approach is not as rich as a full calibration for each new structure. This problem is actually present in many GP works, especially with computationally expensive programs. The potential solution is leveraging surrogate models to approximate the fitness evaluation in GP [41, 42].

5.4. Implications for Complex Systems Modeling Practice. Retroduction—teasing out the complex interaction of real entities and mechanisms that brings about a concrete phenomenon—is challenging. Complex systems models (CSMs) that attempt this are often charged with being arbitrary and/or absolutist in their conceptions of reality. Structural calibration avoids this by looking across a wide multiplicity of models that retain the base elements of mechanism-based theory. Complex systems modelers, who use formal models to help explain concrete phenomena, should use structural calibration as part of their standard modeling practices, in the same way that data-driven modelers, who use formal models to explain variance in patterns of data, consider term selection.

However, automated structural calibration is a major enterprise. It requires complex systems modelers to (i) think more about ontology—what are the base elements of theory that are candidates for inclusion in any model? and (ii) formally describe entities and mechanisms in a consistent way that allows them to be recombined together in meaningful ways. To help, we have developed an open-source

model discovery framework for CSM developers that forces an ontological focus and provides an underlying formal language that is amenable to automatic structural calibration.

This work and the underlying model discovery process contribute to the needs for standards and modeling practices in the ABS community. Collins et al. [43] pointed out that as ABS matures, with many simulation software tools (like Netlogo [44] and Repast Symphony [45]), potential standards and protocols will be needed and proposed. For example, Grimm et al. [46] designed the ODD (Overview, Design concepts, and Details) protocol as a generic and structured template to describe agent-based models for better communication and replicability. Another example is the UML (Unified Modeling Language) to develop and document agent-based models [47–49]. There are also many discussions and proposals about different aspects of the ABS development process. A recently proposed software architecture [50], namely MBSSM (Mechanism-Based Social System Modelling), is designed based on a middle-range theory approach to express individual social theory mechanisms in a unified way. Following this line of thought, our model discovery process can be addressed as a protocol concerned with structural uncertainty. It is common to decide on a model structure and then calibrate the parameters within the model to capture parameter uncertainty. Our work takes this further and addresses the between-model uncertainty of structural assumptions. More importantly, we demonstrated the feasibility of automated model structure discovery by embracing both theory and empirical data. Exploring different model structures is not only valuable for theory testing but also can contribute to theory exploration.

6. Conclusion

Here, we have presented a novel method which utilizes genetic programming techniques to discover new and adapted behavioral theories from models of complex social systems. Using a case study of a social role theory to inform mechanism-based model of population-level alcohol use, we have demonstrated that our approach can find new, theoretically meaningful and interpretable mechanisms which drive population alcohol use in a complex systems model. It would take a human modeler an infeasible amount of time to manually construct a multiplicity of different variations of the mechanisms. The GP method assists in efficient screening of mechanism variants. This screening process can be important, since different realizations of a mechanism can produce qualitatively different model outputs [51]. The novel models generated by the GP method offer a better fit to alcohol consumption data than a reference model, which was constructed by a human modeler representing one possible interpretation of the mechanisms of role theory. Our approach is flexible and can be easily extended to complex systems models that are seeking to explain other social phenomena. Our method also offers novel directions for future knowledge discovery and social theory development, based on the fusion of data-driven and theory-driven methods.

A key part of realist explanation is comparison and integration across multiple theories [4]. While our existing example is limited to the building blocks defined in social role theory, there is no reason why building blocks relating to other theories cannot be defined. However, integration of these wider building blocks, such that they can be exploited by machine learning, will require a common language for expression of the theories. We see middle-range theory [52] and its realization in the so-called Coleman Boat [53], or other micro-macro schemes, as a potentially useful template for formal descriptions of theory and their translation, via a model discovery framework, into integrated simulation models. Future work will aim to incorporate additional theories and to generate novel combinations of multiple theories.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Authors' Contributions

TMV designed the primitives and implemented the model discovery process. AB, CB, and RCP designed the roles model, which was implemented by HB. CB implemented the microsimulation. AN and CP contributed to mechanism development. TMV, AB, MS, and RCP designed the model discovery process. PS aided in the interpretation of the findings. RCP led the research and its conceptual underpinnings. TMV, CB, RCP, MS, and AN wrote the paper.

Acknowledgments

Supplementary Materials

Supplementary Material A—the evolution of the grammar contains three iterations of the grammar with change logs. Supplementary Material B—the structures on the Pareto front and corresponding theoretical credibility. (*Supplementary Materials*)

References

[1] J. M. Epstein, "Agent-based computational models and generative social science," *Complexity*, vol. 4, no. 5, pp. 41–60, 1999.

[2] P. Hedstrom, "Dissecting the social: on the principles of analytical sociology," vol. 34, 2005.

[3] T. Lawson, *Economics and Reality*, Routledge, Abingdon, UK, 1997.

[4] B. Danermark, M. Ekström, L. Jakobsen, and J. Ch. Karlsson, *Explaining Society: Critical Realism in the Social Sciences*, Routledge, Abingdon, UK, 2002.

[5] R. Boero and F. Squazzoni, "Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science," *Journal of Artificial Societies and Social Simulation*, vol. 8, 2005.

[6] F. J. León-Medina, "Analytical sociology and agent-based modeling: is generative sufficiency sufficient?" *Sociological Theory*, vol. 35, no. 3, pp. 157–178, 2017.

[7] D. Byrne and G. Callaghan, *Complexity Theory and the Social Sciences: The State of the Art*, Routledge, Abingdon, UK, 2013.

[8] C. Gunaratne and I. Garibay, "Alternate social theory discovery using genetic programming: towards better understanding the artificial anasazi," in *Proceedings of the GECCO 2017 Genetic and Evolutionary Computation Conference*, ACM Press, Berlin, Germany, pp. 115–122, 2017.

[9] C. Gunaratne and I. Garibay, "Evolutionary model discovery of factors for farm selection by the artificial anasazi," in *Proceedings of the International Conference Computer Science and Engineering*, p. 27, ACM, New York, NY, USA, 2017.

[10] T. M. Vu, C. Probst, J. M. Epstein, A. Brennan, M. Strong, and R. C. Purshouse, "Toward inverse generative social science using multi-objective genetic programming," in *Proceedings of the GECCO 2019 Genetic and Evolutionary Computation Conference*, Association for Computing Machinery, Prague, Czech Republic, pp. 1356–1363, 2019.

[11] V. A. Smith, "Evolving an agent-based model to probe behavioral rules in flocks of cowbirds," *ALIFE*, vol. 2008, pp. 561–568, 2008.

[12] J. Zhong, L. Luo, W. Cai, and M. Lees, "Automatic rule identification for agent-based crowd models through gene expression programming," *AAMAS*, vol. 2014, pp. 1125–1132, 2014.

[13] R. Poli, W. B. Langdon, N. F. McPhee, and J. R. Koza, *A Field Guide to Genetic Programming*, Lulu Press, Morrisville, NC, USA, 2008.

[14] K. Rodriguez-Vazquez, C. M. Fonseca, and P. J. Fleming, "Identifying the structure of nonlinear dynamic systems using multiobjective genetic programming," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 34, no. 4, pp. 531–545, 2004.

[15] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, Cambridge, MA, USA, 1992.

[16] R. I. McKay, N. X. Hoai, P. A. Whigham, Y. Shan, and M. O'Neill, "Grammar-based genetic programming: a survey," *Genetic Programming and Evolvable Machines*, vol. 11, no. 3–4, pp. 365–396, 2010.

[17] B. J. Biddle, *Role Theory: Expectations, Identities, and Behaviors*, Academic Press, Cambridge, MA, USA, 1979.

[18] R. A. Knibbe, M. J. Drop, and A. Muytjens, "Correlates of stages in the progression from everyday drinking to problem drinking," *Social Science & Medicine*, vol. 24, no. 5, pp. 463–473, 1987.

[19] S. Kuntsche, R. A. Knibbe, and G. Gmel, "Social roles and alcohol consumption: a study of 10 industrialised countries," *Social Science & Medicine*, vol. 68, no. 7, pp. 1263–1270, 2009.

[20] R. W. Wilsnack and R. Cheloha, "Women's roles and problem drinking across the lifespan," *Social Problems*, vol. 34, no. 3, pp. 231–248, 1987.

[21] K. Yamaguchi and D. B. Kandel, "On the resolution of role incompatibility: a life event history analysis of family roles and marijuana use," *American Journal of Sociology*, vol. 90, no. 6, pp. 1284–1325, 1985.

[22] J. G. Bachman, P. M. O'Malley, J. E. Schulenberg, L. D. Johnston, A. L. Bryant, and A. C. Merline, *The Decline of Substance Use in Young Adulthood: Changes in Social Activities, Roles, and Beliefs*, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, USA, 2002.

[23] M. R. Lee, L. Chassin, and D. MacKinnon, "The effect of marriage on young adult heavy drinking and its mediators: results from two methods of adjusting for selection into marriage," *Psychology of Addictive Behaviors*, vol. 24, no. 4, pp. 712–718, 2010.

[24] U.S. Department of Health and Human Services, *Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality*, National Survey on Drug Use and Health (NSDUH), New York, NY, USA, 1979.

[25] University of Michigan, *Survey Research Center, Panel Study of Income Dynamics (PSID): Main Interview*, University of Michigan, Ann Arbo, MI, USA, 2018.

[26] S. Manson, J. Schroeder, D. Van Riper, and S. Ruggles, *IPUMS National Historical Geographic Information System: Version 14.0 [Database]*, IPUMS, Minneapolis, MN, USA, 2019.

[27] R. Lovelace and M. Dumont, *Spatial Microsimulation with R*, Chapman and Hall/CRC, New York, NY, USA, 2016.

[28] S. Ruggles, S. Flood, R. Goeken et al., *IPUMS USA: Version 9.0 [dataset]*, IPUMS, Minneapolis, MN, USA, 2019.

[29] Centers for Disease Control and Prevention, National Center for Health Statistics, "Compressed mortality file, 1979–1998. CDC WONDER online database, compiled from compressed mortality file CMF 1968–1988, series 20, no. 2A, 2000 and CMF 1989–1998, series 20, no. 2E," 2003.

[30] Centers for Disease Control and Prevention, National Center for Health Statistics, "Underlying cause of death, 1999–2017 on CDC WONDER online database, released december, 2018. Data are from the multiple cause of death files, 1999–2017, as compiled from data provided by the 57 vital statistics jurisdictions through the vital statistics cooperative program," 1999.

[31] N. Collier and M. North, "Parallel agent-based simulation with repast for high performance computing," *Simulation*, vol. 89, no. 10, pp. 1215–1235, 2013.

[32] R. Carnell, "Lhs R package: latin hypercube samples," 2019.

[33] M. Stein, "Large sample properties of simulations using Latin hypercube sampling," *Technometrics*, vol. 29, no. 2, pp. 143–151, 1987.

[34] M. Fenton, J. McDermott, D. Fagan, S. Forstenlechner, M. O'Neill, and E. Hemberg, "Pony GE2: grammatical evolution in python," *Genetic and Evolutionary Computation Conference*, vol. 17, 2017.

[35] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 2, pp. 182–197, 2002.

[36] L. Gell, G. Bühringer, J. McLeod et al., *What Determines Harm from Addictive Substances and Behaviours?*, Oxford University Press, Oxford, UK, 2016.

[37] J. Staff, K. M. Greene, J. L. Maggs, and I. Schoon, "Family transitions and changes in drinking from adolescence through mid-life," *Addiction*, vol. 109, no. 2, pp. 227–236, 2014.

[38] M. R. Lee, L. Chassin, and D. P. MacKinnon, "Role transitions and young adult maturing out of heavy drinking: evidence for larger effects of marriage among more severe premarriage problem drinkers," *Alcoholism: Clinical and Experimental Research*, vol. 39, no. 6, pp. 1064–1074, 2015.

[39] M. L. Cooper, "Motivations for alcohol use among adolescents: development and validation of a four-factor model," *Psychological Assessment*, vol. 6, no. 2, pp. 117–128, 1994.

[40] M. Nicolau and A. Agapitos, "Understanding grammatical evolution: grammar design," in *Handbook of Grammatical Evolution*, C. Ryan, M. O'Neill, and J. Collins, Eds., pp. 23–53, Springer International Publishing, Berlin, Germany, 2018.

[41] T. Hildebrandt and J. Branke, "On using surrogates with genetic programming," *Evolutionary Computation*, vol. 23, no. 3, pp. 343–367, 2015.

[42] M. Zaefferer, J. Stork, O. Flasch, and T. Bartz-Beielstein, "Linear combination of distance measures for surrogate models in genetic programming," in *Parallel Problem Solving from Nature–PPSN XV*, A. Auger, C. M. Fonseca, N. Lourenço, P. Machado, L. Paquete, and D. Whitley, Eds., pp. 220–231, Springer International Publishing, Berlin, Germany, 2018.

[43] A. Collins, M. Petty, D. Vernon-Bido, and S. Sherfey, "A call to arms: standards for agent-based modeling and simulation," *Journal of Artificial Societies and Social Simulation Index*, vol. 18, p. 12, 2015.

[44] U. Wilnesky, *NetLogo, Center of Connected Learning and Computer-Based Modeling*, Norwestern University, Evanston, IL, USA, 1999.

[45] M. J. North, N. T. Collier, J. Ozik et al., "Complex adaptive systems modeling with repast simphony, complex adapt," *System Model*, vol. 1, p. 3, 2013.

[46] V. Grimm, U. Berger, D. L. DeAngelis, J. G. Polhill, J. Giske, and S. F. Railsback, "The ODD protocol: a review and first update," *Ecological Modelling*, vol. 221, no. 23, pp. 2760–2768, 2010.

[47] B. Bauer, J. P. Müller, and J. Odell, "Agent uml: a formalism for specifying multiagent software systems," *International Journal of Software Engineering and Knowledge Engineering*, vol. 11, no. 03, pp. 207–230, 2001.

[48] H. Bersini, "UML for ABM," *Journal of Artificial Societies and Social Simulation Index*, vol. 15, p. 9, 2011.

[49] T. M. Vu, C. Wagner, and P.-O. Siebers, "ABOOMS: overcoming the hurdles of continuous-time public goods games with a simulation-based approach," *Journal of Artificial Societies and Social Simulation Index*, vol. 22, p. 7, 2019.

[50] T. M. Vu, C. Probst, A. Nielsen et al., "A software architecture for mechanism-based social systems modelling in agent-based simulation models," *Journal of Artificial Societies and Social Simulation*, vol. 23, p. 1, 2020.

[51] H. Muelder and T. Filatova, "One theory - many formalizations: testing different code implementations of the theory of planned behaviour in energy agent-based models," *Journal of Artificial Societies and Social Simulation Index*, vol. 21, p. 5, 2018.

[52] R. K. Merton, *On Theoretical Sociology*, Free Press, New York, NY, USA, 1967.

[53] J. S. Coleman, "Social theory, social research, and a theory of action," *American Journal of Sociology*, vol. 91, no. 6, pp. 1309–1335, 1986.