

Toward inverse generative social science using multi-objective genetic programming

Tuong Manh Vu
University of Sheffield
Sheffield, UK
t.vu@sheffield.ac.uk

Charlotte Probst
Centre for Addiction & Mental Health
Toronto, Canada
mariecharlotte.probst@gmail.com

Joshua M. Epstein*
New York University
New York City, USA
je65@nyu.edu

Alan Brennan
University of Sheffield
Sheffield, UK
a.brennan@sheffield.ac.uk

Mark Strong
University of Sheffield
Sheffield, UK
m.strong@sheffield.ac.uk

Robin C. Purshouse†
University of Sheffield
Sheffield, UK
r.purshouse@sheffield.ac.uk

ABSTRACT

Generative mechanism-based models of social systems, such as those represented by agent-based simulations, require that intra-agent equations (or rules) be specified. However there are often many different choices available for specifying these equations, which can still be interpreted as falling within a particular class of mechanisms. Whilst it is important for a generative model to reproduce historically observed dynamics, it is also important for the model to be theoretically enlightening. Genetic programs (our own included) often produce concatenations that are highly predictive but are complex and hard to interpret theoretically. Here, we develop a new method – based on multi-objective genetic programming – for automating the exploration of both objectives simultaneously. We demonstrate the method by evolving the equations for an existing agent-based simulation of alcohol use behaviors based on social norms theory, the initial model structure for which was developed by a team of human modelers. We discover a trade-off between empirical fit and theoretical interpretability that offers insight into the social norms processes that influence the change and stasis in alcohol use behaviors over time.

CCS CONCEPTS

• **Computing methodologies** → **Genetic programming**; **Modeling methodologies**; **Model verification and validation**; • **Applied computing** → **Sociology**.

KEYWORDS

generative social science, multi-objective optimization, genetic programming

*Joshua Epstein is also an External Professor at the Santa Fe Institute

†Corresponding author: Robin Purshouse, Department of Automatic Control & Systems Engineering, University of Sheffield, Mappin Street, Sheffield, S1 3JD, UK

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

GECCO '19, July 13–17, 2019, Prague, Czech Republic

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6111-8/19/07.

<https://doi.org/10.1145/3321707.3321840>

ACM Reference Format:

Tuong Manh Vu, Charlotte Probst, Joshua M. Epstein, Alan Brennan, Mark Strong, and Robin C. Purshouse. 2019. Toward inverse generative social science using multi-objective genetic programming. In *Genetic and Evolutionary Computation Conference (GECCO '19)*, July 13–17, 2019, Prague, Czech Republic. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3321707.3321840>

1 INTRODUCTION

1.1 Inverse generative social science

In describing *Inverse Generative Social Science*, Epstein writes that, “Artificial intelligence and machine learning are displacing humans, but not explaining them. Machines can crush humans at chess, but do not illuminate how humans play chess. When asked how he came up with a winning brilliancy against IBM’s Deep Blue, Gary Kasparov answered simply: ‘it smelled right.’ While it has not been used in this way, machine learning can help explain human behavior, even moving our modeling of complex social systems into a new epoch.” [7] The agent-based model (ABM) is the principal scientific instrument for understanding how individual behaviors and interactions, the micro-world, generates change and stasis in macroscopic social regularities. So far, agents have been iterated forward to generate such explananda as settlement patterns, scaling laws, epidemic dynamics, and many other phenomena [6]. But these are all examples of the forward problem: we design agents and grow the target phenomenon. The motto of generative social science is: “If you didn’t grow it, you didn’t explain it.” [5] But there may be many ways to grow it! How do we find ‘all’ the non-trivial generators? This is inverse generative social science – agent architectures as model outputs not model inputs – and machine learning can enable it. Our application of multi-objective genetic programming in the present paper represents a starting point for building the tools needed to perform the model discovery process of inverse generative social science. Our focus is the internal rules that describe the dispositional dynamics of the agents, since these constitute a pillar of the agent-based generative approach. However we are clear that a comprehensive model discovery process would also include identification of social contact structures such as networks.

1.2 Related works

Evolutionary computation forms one of the main classes of method for calibration of an agent-based model's *parameters*, given a pre-defined model structure [23]. These methods emulate natural selection processes in order to find the numerical parameters that optimize an ABM's goodness-of-fit metric when comparing model outputs to empirical observations. However it is only relatively recently, in a handful of studies, that evolutionary computing has also been used to search for agent-based model *structures* – the agents' internal rules and structuring computational architectures. In an early study, Smith used a genetic algorithm to evolve the rules in a classifier in order to reproduce the observed social assortativity of birds [22]. More recently, Zhong and colleagues used gene expression programming to optimize the structure of a reward function used by agents to evaluate behavioral choices, such that the ABM could better reproduce empirically observed crowd behaviors [28]. Later, Gunaratne & Garibay used genetic programming to evolve agents' farm selection rules to identify new model structures for a NetLogo implementation of the seminal Artificial Anasazi ABM in order to reproduce the archeological population demography of Long House Valley, Arizona [9]. In all three cases, the focus for structural calibration was on aspects of agent decision-making, rather than any other parts of the model.

The multi-agent system (MAS) community faces a related – yet distinct – problem to ABM structural calibration, where the engineered system is seeking to fulfill an overall desired objective (rather than replicate empirical observations). Whilst the application is most emphatically not inverse generative social science, the evolutionary computing methods used in these studies, again focusing on agent decision-making processes, are very relevant to the generic requirement for model discovery. Junges & Klügl used genetic programming to find decision tree representations of agent decision-making that minimized collisions in a pedestrian evacuation scenario [13]. van Berkel and colleagues developed infrastructure for automatic discovery of behavioral algorithms for MAS, in which grammatical evolution was used to find optimal algorithm structures for solving five types of multi-agent problem, based on defined primitives implemented in the NetLogo ABM software [25]. Husselmann and colleagues extended a domain-specific language known as SOL to enable parallel search over evolved sets of abstract syntax trees, applying the method to solve two multi-agent problems [11, 12]. In both latter cases, researchers have been particularly sensitive to the computational burden of performing structural optimization in agent-based contexts, developing enabling, although perhaps rather esoteric, infrastructure to help meet these needs.

1.3 Aim of the present study

This paper seeks to demonstrate how multi-objective genetic programming can be used as a tool for inverse generative social science. This aim is supported by using the case study of an existing individual-level simulation model developed to explain trends in patterns of alcohol use in the United States. In Section 2, this model is introduced in sufficient detail for the subsequent manipulation of its structure to be understood. Then, in Section 3, an overall process of model discovery is introduced, with particular focus on

the evolutionary computing tools used to search through the space of model structures. Results of the model discovery process for the case study are provided in Section 4. The findings are discussed in detail in Section 5. The field of inverse generative social science is in its infancy and so the paper closes with a plethora of directions for future research on this exciting topic.

2 AN INDIVIDUAL-LEVEL MODEL OF SOCIAL NORMS

2.1 Overview

The present study makes use of a mathematical model which aims to explain the change and stasis in patterns of alcohol use within the United States over the past four decades. The model is based on theoretical concepts from social norms theory [4], in which the modeler has used the theory abductively as a framework for interpretation of the temporal changes in alcohol use patterns in American society [2]. The conceptual design of the model uses the *analytical sociology* framework [18], in which the sequence of mechanisms that generate the dynamics of macro-level patterns involve, for every mechanism, the dispositions and behaviors of individuals. Empirical-based falsification is also a key part of the framework – with mechanism-based explanations only remaining as candidates if they are able to sufficiently reproduce observations.

2.2 The phenomenon to be explained

The model attempts to reproduce population-level patterns of alcohol use, in terms of prevalence of use in the last 12 months, average frequency of use (in terms of average number of drinking days in a month), and average quantity of use (in terms of average grams of ethanol consumed per day). Empirical data on these measures comes from the National Survey on Drug Use and Health (NSDUH, until 2002 called the National Household Survey on Drug Abuse), which is a nationally representative, cross-sectional survey of the noninstitutionalized civilian population of the US aged 12 or older [24]. The survey has been conducted annually since 1990, and before then in the years 1979, 1982, 1985 and 1988. The observed patterns of alcohol use are shown, separately for males and females, in Figure 1. Whilst alcohol use prevalence is observed to decline over the period 1979 to 2016, a corresponding initial decline in drinking frequency is observed to reverse its trend in the mid-1990s. Also of interest, since the mid-2000s, average drinking quantities per day have been observed to be decreasing in males but to be stable or slightly increasing in females.

2.3 Model description

While we will exercise the modeling approach in the context of alcohol, it will be useful to step back and briefly discuss the overarching idea, as it applies to social norms in general. As per Einstein's famous adage, we want the simplest model capturing the main drivers; therefore when developing the social norms model, the modelers in question have aimed for parsimony. The dependent variable is the agent's disposition to engage in the behavior of interest (e.g., drinking, smoking, consuming drugs). There are but four independent variables: payoff, autonomy, and injunctive and descriptive norms. These all range from zero to one and can vary

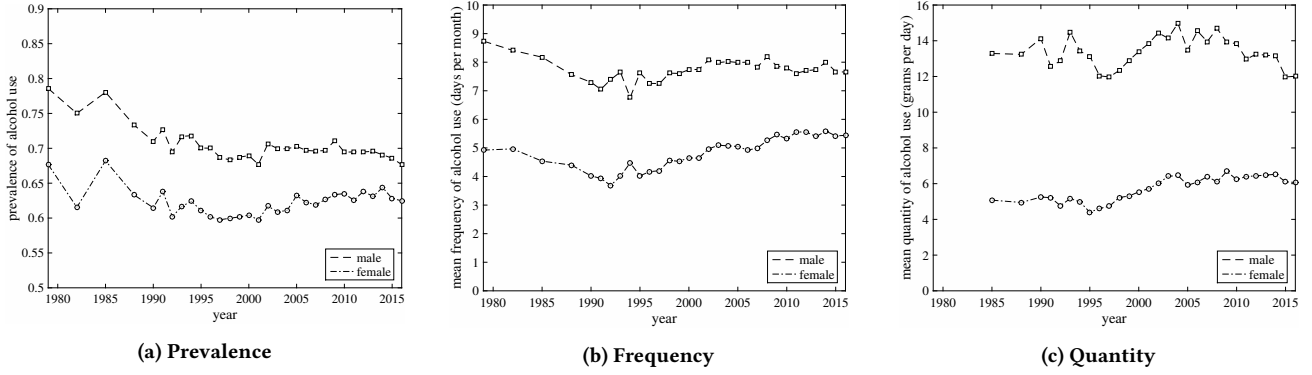


Figure 1: Empirical observations used as targets in the model discovery process: alcohol use dynamics in the United States, 1979–2016

across agents, by gender, age and other heterogeneous attributes, and also change over time as affected by agent behavior. In the health context, payoff is hedonic and could be the neuro-biological satisfaction derived from consumption. The agent also makes an appraisal of the social opprobrium associated with the behavior. This is the injunctive norm, capturing the agent’s perceived rules and expectations regarding acceptable and unacceptable drinking behavior. The descriptive norm is the agent’s empirical estimate of the behavior’s prevalence in a relevant reference group. In this case, it is the perceived fraction of individuals who drink and the perceived amount consumed on a drinking occasion. The perception of descriptive norms is imperfect as it is biased by the agent’s own drinking habits.

The crux of the model is autonomy – the agent’s disregard for social norms. An agent with autonomy of one does not care about society and is driven purely by her individual hedonic payoff. At the other extreme is the agent who is a ‘slave to fashion’ and exhibits no autonomy whatsoever. She is willing to sacrifice individual pleasure for social conformity and acceptance. Disposition is the convex combination of individual and social forces, as autonomy ranges from zero to one. How then to represent these social forces, the influence of the injunctive and descriptive norms? Because all variables are bounded to the unit interval, an initial hypothesis postulated by the modelers is that the social force term could be the geometric mean of these factors. The exact structure – how these concepts work together – remains tentative. It is hoped that the process of model discovery can elucidate new expressions to stimulate further theorizing about the model, but in a way that is empirically grounded.

For context, it should be noted that the social norms model contains other mechanisms beside that which drives agents’ dispositions. While the agent’s behavior is – to a varying degree – impacted by the social norms, the inverse is also true. In taking decisions about individual drinking, the agents transform the descriptive norms that are perceived by their peers. Furthermore, excessive drinking can become less acceptable over time when it is so prevalent that it causes perceptible social, material or physical harm to society and its members. Via the perception of descriptive

and injunctive norms on the one hand and the collective transformation of social norms on the other, individual agents form a dynamic agent society. The emerging population-level drinking behaviors are therefore a result of individual-level interactions between global social norms and individual desire to drink.

2.4 Implementation

The social norms model follows an object-oriented design using the mediator design pattern. The model is implemented in C++ using Repast HPC libraries [1]. For the present study, the model was coarsely parallelised across 36 cores of an Intel i9 9980XE processor by assigning separate sub-populations to each core.

3 MODEL DISCOVERY METHODS

3.1 Overview

A flow chart depicting the overall process of model discovery is shown in Figure 2. The process both extends and formalises the recent approach described by Gunaratne & Garibay [9]. The different steps in the process are discussed here, making reference to the case study model introduced previously.

In **Step 1**, we make explicit the abductive reasoning process of developing a mechanism-based mathematical model – at present there is no library of context-free model components or “building blocks” [17] that can be drawn upon to automate this reasoning stage. In summary, we expect to be starting with a specific model developed by a human modeler.

The social norms model described in Section 2 comprises a number of entities and mechanisms. Here, we focus on the *situational mechanism* [10] by which an individual’s disposition to drink (defined by the agent property disposition) is affected by perceptions of norms (properties descriptive and injunctive), in combination with his or her desire to disregard norms (property autonomy) and beliefs about the value of drinking (property payoff). The original equation that defines this mechanism is given in Equation 1.

$$\text{disposition} = (\text{autonomy} * \text{payoff}) + (1 - \text{autonomy}) * \text{sqrt}(\text{injunctive} * \text{descriptive}) \quad (1)$$

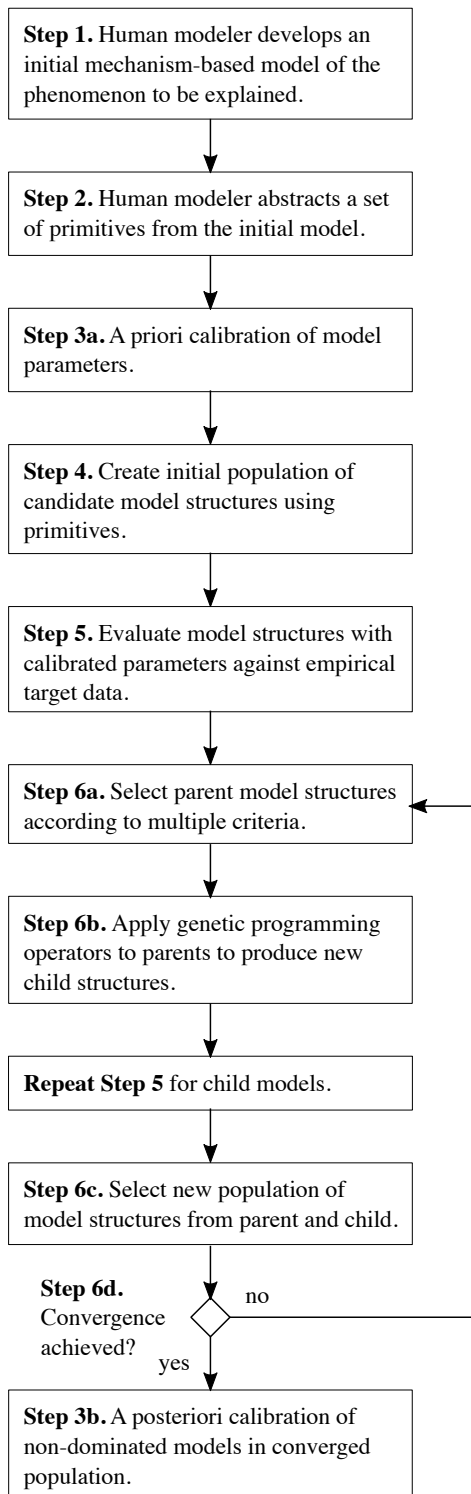


Figure 2: Model discovery process

Note that agents are heterogeneous in their autonomy and in their appraisals of injunctive and descriptive norms, as well as by gender, age, and other factors that may also be dynamic. The above equation describes the single explanatory architecture that induces disposition in every agent, but agent dispositions have the capacity to be heterogeneous due to the heterogeneity in their properties.

3.2 Structural primitives

In Step 2, we work with the modeler (or relevant documentation) to develop a set of primitives that can be used as the building blocks in the creation of new model structures. In our case study, the primitives relate simply to elements of the situational mechanism given in Equation 1. The modeler in this case was also uncertain about whether descriptive and injunctive norms should be averaged arithmetically or geometrically, so we included both operators as functions in the set of primitives. We also included a composite function complement rather than introduce constants such as 1. The full set of primitives used in the model discovery process is shown in Table 1.

3.3 Calibration of model parameters

The full social norms model contains 27 parameters that must be specified prior to running the model – these relate both to effect sizes in mechanisms and also hyperparameters that define the heterogeneity of properties (such as autonomy) in the population of agents. It is conventional practice, in empirically grounded studies, to calibrate such parameters using available evidence [23]. If the calibration procedure were cheap, we would calibrate the parameters afresh for each and every model structure proposed during the model discovery process. However, the computational cost of calibration makes this infeasible, and we must be more circumspect in how often we choose to run a parameter calibration for a given structure. In the present process, we split parameter calibration into two stages. In Step 3a a single parameter calibration is performed using a baseline model structure (i.e. Equation 1). This set of parameters is then used for all other model structures during the model discovery process. Once the process has converged, a further round of calibration is performed in Step 3b for each model in the set of structures that are most promising according to two criteria described below. Clearly, some promising model structures will be penalised by this approach and we return to this topic in Section 5.

For the case study, we employ a Latin hypercube space-filling design to sample 5,000 parameter sets from the joint prior distribution that represents judgements about plausible values. We then evaluate the model at these parameter sets and assess the model output against empirical observations using a goodness-of-fit metric, and select the best-fitting parameterisation. In each case, the social norms model is initialised in 1979 with 1,000 agents (sampled from a microsynthesis of the US population based on 1980 census data) and run forward in time for 15 years. The goodness-of-fit metric quantifies the distance between the model outputs for drinking prevalence, frequency and quantity (split by male and female sub-populations) and their corresponding empirically observed values (see Figure 1).

This goodness-of-fit metric also forms one of the two objective functions used in the model structure discovery process step of the

Table 1: Model primitives derived from the social norms model

Type	Name	Arity	Description
Terminals	injunctive	0	Perceived injunctive norm
	descriptive	0	Perceived descriptive norm
	payoff	0	Positive utility derived from drinking alcohol
	autonomy	0	Tendency to ignore social norms
Functions	+	2	Adds two quantities, i.e. $x + y$
	-	2	Subtracts two quantities, i.e. $x - y$
	*	2	Multiplies two quantities, i.e. xy
	mean	2	Takes the arithmetic mean of two quantities, i.e. $(x + y)/2$
	root	1	Takes the square root of a quantity, i.e. \sqrt{x}
	complement	1	Takes the complement of a quantity, i.e. $1 - x$

calibration process. The objective function for model structure \mathbf{x} is an *implausibility* metric [27] that measures the error between m simulated outputs y_m^* and equivalent empirical target data y_m over temporal observations k defined by:

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

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 for output m , which is taken as 10% of the possible output range for each output. ‘Model discrepancy’ is the error in a model output that arises because the model is not a perfect representation of reality.

3.4 Evolutionary search using genetic programming

The evolutionary computing approach requires specification of: (i) a representation of a candidate solution; (ii) an objective function for evaluating each candidate; (iii) selection mechanisms for emulating survival-of-the-fittest mechanisms; (iv) variation operators for manipulating existing candidates to form new candidates; and (v) a convergence measure that can act as a stopping criterion. Each of these aspects forms part of the model discovery process and is discussed in more detail below.

In order to represent new models (essentially new right-hand sides of the situational mechanism given in Equation 1) we adopt the syntax tree structure used in conventional genetic programming [14]. The tree represents the terminals and functions abstracted during Step 2 (shown in Table 1). In **Step 4** an initial population of trees (representing new model structures) is then created – here we adopt the *ramped half-and-half* tree-building method developed by Koza [14, 15]. Next, in **Step 5**, we simulate the new model structures and compare them to empirical target data using Equation 2.

For inverse generative social science, goodness-of-fit is not the only criterion by which the worth of a model should be measured. Here, we consider one of the other critically important criteria – interpretability. A substantive, although not only, part of interpretability is the number of terms in the model’s equations, since there are cognitive limits on how a large a structure can be comprehended. However, for the present time we use this complexity

measure as the second criterion. The choice of a complexity measure is, of course, quite common in model selection more generally and also of relevance to the issue of *bloat* in the specific context of genetic programming [20]. We calculate interpretability therefore as the number of nodes in the tree:

$$z_2 = \text{nodes}(\mathbf{x}), \quad (3)$$

where $\text{nodes}(\cdot)$ calculates the number of nodes in the tree encoding model structure \mathbf{x} .

Step 6 forms the heart of the evolutionary approach. Objectives $\mathbf{z} = [z_1, z_2]$ are to be minimized simultaneously. In the case study we have used the selection mechanisms of the *NSGA-II* optimizer [3]. This algorithm is suitable for unconstrained bi-objective problems and aims to develop an even, sample-based representation of the Pareto front representing the anticipated trade-off between goodness-of-fit and interpretability. *NSGA-II*’s conventional variation operators are replaced with equivalents for tree-based genetic programming [14, 15].

The resulting multi-objective genetic programming (MOGP) algorithm is run until convergence is achieved or the resources for optimization have been exhausted (**Step 7**). In this study, we monitored convergence using the hypervolume indicator, which is a popular performance metric for multi-objective problems [8].

3.5 Implementation details

The multi-objective genetic programming algorithm was implemented using existing components from the *ECJ* toolkit [16]. To evaluate a new model structure generated by the algorithm, additional scripting was necessary to: (i) modify the C++ source code of the situational mechanism embedded in the agents decision making; (ii) re-compile the overall Repast HPC model; (iii) execute the simulation; and (iv) compare simulated population-level outputs against empirical targets. The population of the MOGP was set at 450 and it was allowed to run for 50 generations, during which convergence was monitored. During testing, little progress was identified in the hypervolume measure following generation 35.

4 RESULTS

The optimizer ran for 50 generations, by which time stability was observed in the hypervolume measure of convergence. The final

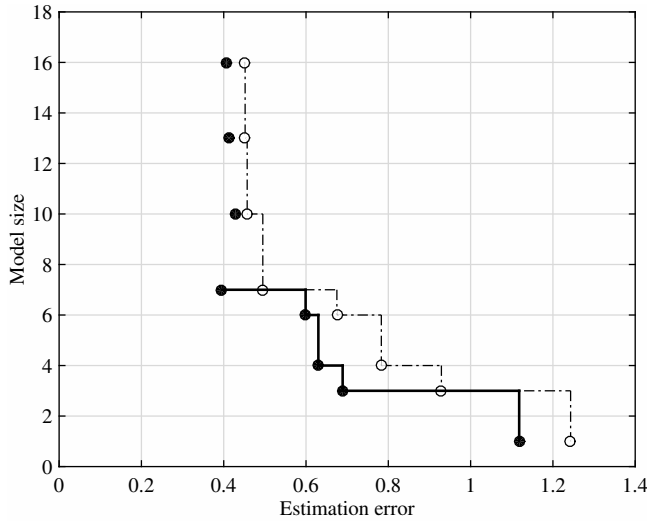


Figure 3: Pareto front of promising model structures:
 ○ uncalibrated models; ● calibrated models; --- uncalibrated Pareto front; — calibrated Pareto front

population of the optimizer was filtered for non-dominated model structures, of which 8 were identified. The parameters of these models were then re-calibrated for each model in turn. The performance of the models, in terms of both estimation error against target data (before and after calibration) and also structural complexity, is shown in the Pareto plot in Figure 3. The corresponding details of the model structures are shown in Table 2. The time series plots that underpin the estimation errors are shown in Figure 4 for three exemplary model structures.

The *pre-calibration* Pareto front is convex and offers good coverage of the trade-off between estimation accuracy and structural complexity: errors range between approximately 0.4 and 1.2 units of empirical implausibility, with structures ranging in size from 16 nodes to just one node. A so-called “knee” in the trade-off surface is apparent at the seven node model, GP4, where increases in model complexity provide rapidly diminishing returns in terms of improved non-implausibility. As clear from Table 2, model structures become substantially more difficult to interpret for tree sizes in double figures.

The *post-calibration* Pareto front shows reductions in implausibility for all models, as expected. However only relatively small improvements are seen for the three most complex models (GP1-3) in comparison to the less complex models, such that GP1-3 become dominated by the seven node GP4. The remaining models remain non-dominated with respect to each other and comprise the remainder of the post-calibration trade-off surface.

5 DISCUSSION

5.1 Findings

The model discovery process has produced a number of interesting findings. Firstly, of the four possible choices for the single-terminal structure given in Table 1, the MOGP identified disposition = payoff as the non-dominated option. This result can be traced

back to the choice of targets shown in Figure 1 combined with the averaging nature of the goodness-of-fit metric in Equation 2 – whilst patterns of alcohol use in US society are not in stasis, they can be relatively well approximated *in the long-run over the calibration window* by a straight line. Since payoff is initialised using agent drinking at baseline, the notion that “past behavior predicts future behavior” is sufficient to offer an explanation within almost one unit of implausibility (although, from the plots in Figure 4, it is apparent that this model, GP8, has particular issues representing observed drinking prevalences).

Secondly, payoff is more implausible than other model structures, including those involving normative concepts as components. Technically, we have not tested an orthodox utility-maximizing agent with prices and budget constraints. Nonetheless, it is intriguing that the closest relatives to homo economicus – the solutions that include payoff only – rank at the bottom of the rules evolved by the MOGP. Norms, typically absent from standard economic models, loom large in the better-fitting variants. In particular, concepts of autonomy and descriptive norms are needed to reduce the estimation error of the model below 0.5 units of implausibility.

Thirdly, we see extensive use of square-root operators in better-fitting models identified by the MOGP; and we also observe little benefit when calibrating the parameters of these models post-discovery. Recalling that the MOGP was unable to re-calibrate any of the structures that it identified *during* the discovery process, our results suggest that the algorithm is implicitly attempting to assign effect modifiers through use of the root function as a means of compensation. It would be interesting to see if this behavior would be lessened if the MOGP were able to perform parameter calibration during the search process.

5.2 Joint structure-parameter model discovery

In the current state-of-the-art process for model discovery – as represented by Figure 2 – parameter calibration only occurs at the start and the end of the process. This is a clear limitation of the approach and a central research challenge for future efforts. Joint structure-parameter calibration can be conceived of as a *bi-level* optimization problem, in which there exists a corresponding parameter calibration problem for every candidate model structure. Methods for solving bi-level optimization problems are also somewhat in their infancy; however there are a number of approaches based on surrogate modeling of fitness functions or best-responses in a Stackelberg game that are worthy of investigation in the cause of inverse generative social science [21].

5.3 Managing uncertainty

Empirically-grounded models of social systems often have to cope with variable-quality empirical evidence with which to calibrate and validate the models – fundamentally, uncertainty abounds and this uncertainty should be represented in the model discovery process and accounted for in conclusions drawn from the findings of the process. For example, our model structure GP4 is seen to dominate the only identified structure that includes injunctive norms, GP1 – but the difference in model fit is only one hundredth of a unit of implausibility. Does this finding present sufficient grounds to falsify injunctive norms as explanatory components of alcohol use patterns

Table 2: Estimated Pareto optimal structures for an agent’s disposition to drink

ID	Estimation error		Model size	Structure
	Uncalibrated	Calibrated		
GP1	0.451	0.407	16	$(\text{payoff} * (\text{autonomy} + \sqrt{((1 - (1 - \sqrt{(\sqrt{\text{descriptive}} + \text{injunctive})))})} + \text{descriptive}))) * \text{payoff}$
GP2	0.452	0.412	13	$((\text{autonomy} + \sqrt{(\sqrt{(\text{autonomy} + \text{descriptive})})} + \text{descriptive})) * \text{payoff} * \text{payoff}$
GP3	0.457	0.427	10	$(\text{payoff} * (\text{autonomy} + \sqrt{(\text{autonomy} + \text{payoff})})) * \text{payoff}$
GP4	0.495	0.395	7	$((\text{autonomy} * \text{payoff}) + \text{payoff}) * \text{payoff}$
GP5	0.675	0.599	6	$(\text{descriptive} + \sqrt{\text{payoff}}) * \text{payoff}$
GP6	0.783	0.630	4	$\text{payoff} * \sqrt{\text{payoff}}$
GP7	0.929	0.689	3	$\text{payoff} * \text{payoff}$
GP8	1.243	1.118	1	payoff

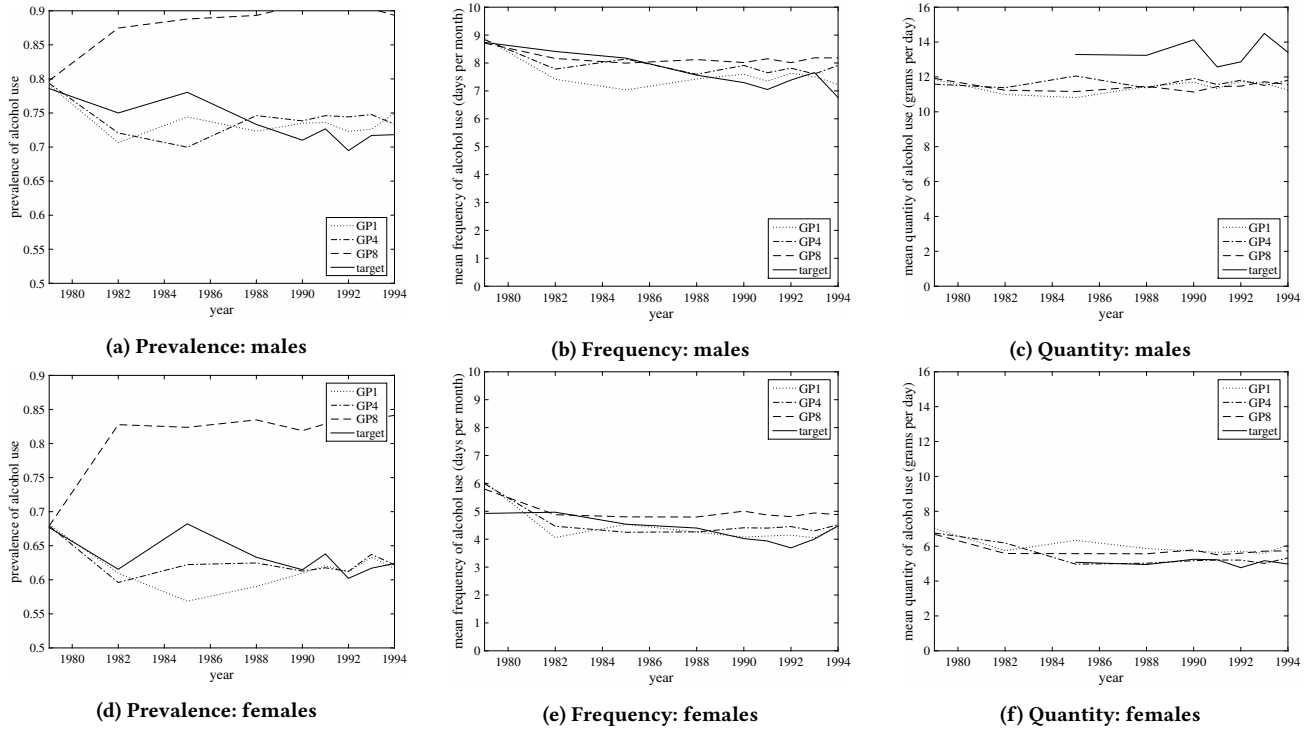


Figure 4: Model time series results for the most complex model (··· GP1), best-fitting model (- · - GP4), and least complex model (- - - GP8), in comparison to target data (—)

in the US during the 1980s and '90s? We would argue not, and that an explicitly probabilistic approach to uncertainty management would offer a more nuanced approach to model falsification.

Bayesian methods offer a suitable approach for managing uncertainty in simulation models, where the class of methods known as *Approximate Bayesian Computation* has been developed to cope with the lack of an explicit likelihood function [26]. At present, these

methods are limited to parameter calibration but have the potential for extension to structural calibration. It is important to recognise that most *evolutionary* approaches to calibration are not probabilistic, although Purshouse et al. [19] presented a quasi-probabilistic evolutionary approach in the context of individual-level model parameter calibration. It remains to be seen how methods such as genetic programming can be used within a probabilistic framework.

5.4 Fully unleashed inverse generative social science

The small number of existing works on inverse generative social science, including the present paper, have been very modest in the degree of model discovery attempted. In all cases, the studies have focused on modifying a single aspect of agent decision-making. They have not attempted to discover wider aspects of the models, such as network structures or alternative types of mechanism (e.g. transformational mechanisms [10]). To fully unleash the power of inverse generative social science, full models should be opened up to the process of model discovery. Further, we should aim to compare and combine model components drawn from *multiple models*, where those models have each been developed from a particular theoretical perspective. Such ambition also raises important questions around model identifiability and the most appropriate level at which model primitives should be specified.

6 CONCLUSION

Inverse generative social science offers new possibilities for the development of empirically-grounded social theory. In this paper, we have developed a first formal process for model discovery that can underpin inverse generative efforts. We have harnessed a multi-objective genetic programming framework for this purpose and demonstrated how it can be used to identify promising new models for social norms-based theorising in a practical application. A multi-objective perspective on model discovery can offer important insights into the trade-off between empirical goodness-of-fit and theoretical interpretability. Meanwhile, consideration of evidential uncertainties remains a key challenge for evolutionary approaches, such as genetic programming, to address.

ACKNOWLEDGMENTS

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. The authors have no conflicts of interest. TMV designed the primitives and implemented the model discovery process; CP developed the social norms model; TMV, MS, AB and RP designed the model discovery process. JME and RCP led the research and its conceptual underpinnings. TMV, CP, JME and RCP wrote the manuscript. The authors would also like to acknowledge support from the wider R01AA024443 'CASCADE' project team.

REFERENCES

- [1] N. Collier and M. North. 2013. Parallel agent-based simulation with Repast for High Performance Computing. *SIMULATION* 89, 10 (2013), 1215–1235. <https://doi.org/10.1177/0037549712462620>
- [2] B. Danermark, M. Ekström, L. Jakobsen, and J. Ch. Karlsson. 2002. *Explaining Society: Critical realism in the social sciences*. Routledge, London.
- [3] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation* 6, 2 (2002), 182–197. <https://doi.org/10.1109/4235.996017>
- [4] C. Elsenbroich and N. Gilbert. 2014. *Modelling Norms*. Springer, Dordrecht. <https://doi.org/10.1007/978-94-007-7052-2>
- [5] J.M. Epstein. 1999. Agent-based computational models and generative social science. *Complexity* 4, 5 (1999), 41–60. [https://doi.org/10.1002/\(SICI\)1099-0526\(199905/06\)4:5<41::AID-CPLX9>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1099-0526(199905/06)4:5<41::AID-CPLX9>3.0.CO;2-F)
- [6] J.M. Epstein. 2012. *Generative Social Science: Studies in agent-based computational modeling*. Princeton University Press, Princeton.
- [7] J.M. Epstein. 2019. Inverse Generative Social Science: What Machine Learning can do for Agent-Based Modeling. In *Social-Behavioral Modeling for Complex Systems*, P.K. Davis, A. O'Mahony, and J. Pfautz (Eds.). Wiley, (in press).
- [8] M. Fleischer. 2003. The measure of Pareto optima applications to multi-objective metaheuristics. *Lecture Notes in Computer Science* 2632 (2003), 519–533.
- [9] C. Gunaratne and I. Garibay. 2017. Alternate social theory discovery using genetic programming: Towards better understanding the artificial Anasazi. *GECCO 2017 - Proceedings of the 2017 Genetic and Evolutionary Computation Conference* (2017), 115–122. <https://doi.org/10.1145/3071178.3071332>
- [10] P. Hedström and R. Swedberg. 1996. Social Mechanisms. *Acta Sociologica* 39, 3 (1996), 281–308. <https://doi.org/10.1177/000169939603900302>
- [11] A.V. Husselmann and K.A. Hawick. 2014. Automatic high performance structural optimisation for agent-based models. In *Proceedings of the 14th International Conference on Software Engineering Research and Practice (SERP'14)*, 1–7.
- [12] A.V. Husselmann, C.J. Scogings, and K.A. Hawick. 2016. Model structure optimisation in lattice-oriented agent-based models. *International Journal of Modelling and Simulation* 36, 4 (2016), 136–144. <https://doi.org/10.1080/02286203.2016.1142276>
- [13] R. Junges and F. Klügl. 2011. Evolution for modeling - A genetic programming framework for SeSAM. *Genetic and Evolutionary Computation Conference, GECCO'11 - Companion Publication* (2011), 551–558. <https://doi.org/10.1145/2001858.2002047>
- [14] J.R. Koza. 1992. *Genetic Programming: On the programming of computers by means of natural selection*. MIT Press, Cambridge, MA.
- [15] J.R. Koza. 1994. *Genetic Programming II: On the programming of computers by means of natural selection*. MIT Press, Cambridge, MA.
- [16] S. Luke. 2017. ECJ then and now. *GECCO 2017 - Proceedings of the Genetic and Evolutionary Computation Conference Companion* (2017), 1223–1230. <https://doi.org/10.1145/3067695.3082467>
- [17] P. Machamer, L. Darden, and C.F. Craver. 2000. Thinking about mechanisms. *Philosophy of Science* 67, 1 (2000), 1–25. <https://doi.org/10.1086/392759>
- [18] G. Manzo. 2014. Data, Generative Models, and Mechanisms: More on the principles of analytical sociology. In *Analytical Sociology: Actions and networks*, G. Manzo (Ed.). Chichester, 1–52. <https://doi.org/10.1002/9781118762707.ch01>
- [19] R.C. Purshouse, A.K. Ally, A. Brennan, D. Moyo, and P. Norman. 2014. Evolutionary parameter estimation for a theory of planned behaviour microsimulation of alcohol consumption dynamics in an English birth cohort 2003 to 2010. *GECCO 2014 - Proceedings of the 2014 Genetic and Evolutionary Computation Conference* (2014), 1159–1166. <https://doi.org/10.1145/2576768.2598239>
- [20] K. Rodríguez-Vázquez, C.M. Fonseca, and P.J. Fleming. 2004. Identifying the structure of nonlinear dynamic systems using multiobjective genetic programming. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, 34, 4 (2004), 531–545. <https://doi.org/10.1109/TSMCA.2004.826299>
- [21] A. Sinha, P. Malo, and K. Deb. 2018. A Review on Bilevel Optimization: From Classical to Evolutionary Approaches and Applications. *IEEE Transactions on Evolutionary Computation* 22, 2 (2018), 276–295. <https://doi.org/10.1109/TEVC.2017.2712906>
- [22] V.A. Smith. 2008. Evolving an agent-based model to probe behavioral rules in flocks of cowbirds. *Artificial Life XI: Proceedings of the 11th International Conference on the Simulation and Synthesis of Living Systems, ALIFE 2008* (2008), 561–568.
- [23] J.C. Thiele, W. Kurth, and V. Grimm. 2014. Facilitating Parameter Estimation and Sensitivity Analysis of Agent-Based Models: A Cookbook Using NetLogo and 'R'. *Journal of Artificial Societies and Social Simulation* 17, 3 (2014), 11. <https://doi.org/10.18564/jasss.2503>
- [24] United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Center for Behavioral Health Statistics and Quality. 2016. National Survey on Drug Use and Health, 2016, 34481-v3. <https://doi.org/10.3886/34481.v3>
- [25] S. van Berkel, D. Turi, A. Pruteanu, and S. Dulman. 2012. Automatic discovery of algorithms for multi-agent systems. *GECCO'12 - Proceedings of the 14th International Conference on Genetic and Evolutionary Computation Companion* (2012), 337–344. <https://doi.org/10.1145/2330784.2330833>
- [26] E. van der Vaart, M.A. Beaumont, A.S.A. Johnston, and R.M. Sibly. 2015. Calibration and evaluation of individual-based models using Approximate Bayesian Computation. *Ecological Modelling* 312 (2015), 182–190. <https://doi.org/10.1016/j.ecolmodel.2015.05.020>
- [27] I. Vernon, M. Goldstein, and R.G. Bower. 2010. Galaxy formation: A Bayesian uncertainty analysis. *Bayesian Analysis* 5, 4 (2010), 619–670. <https://doi.org/10.1214/10-BA524>
- [28] J. Zhong, L. Luo, W. Cai, and M. Lees. 2014. Automatic rule identification for agent-based crowd models through gene expression programming. *13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014 2* (2014), 1125–1132.