Alternate Social Theory Discovery Using Genetic Programming: Towards Better Understanding the Artificial Anasazi

Chathika Gunaratne
University of Central Florida
Orlando, FL
chathika@knights.ucf.edu

Ivan Garibay
University of Central Florida
Orlando, FL
igaribay@ucf.edu

ABSTRACT

A pressing issue with agent-based model (ABM) replicability is the ambiguity behind micro-behavior rules of the agents. In practice, modelers choose between competing theories, each describing separate candidate solutions. Pattern-oriented modeling (POM) and stylized facts matching recommend testing theories against patterns extracted from real-world data. Yet, manually, POM is tedious and prone to human error. In this study, we present a genetic programming strategy to evolve debatable assumptions on agent micro-behaviors. After proper modularization of the candidate micro-behaviors, genetic programming can discover candidate micro-behaviors which reproduce patterns found in real-world data. We illustrate this strategy by evolving the decision tree representing the farm-seeking strategy of agents in the Artificial Anasazi ABM. Through evolutionary theory discovery, we obtain multiple candidate decision trees for farm-seeking which fit the archaeological data better than the calibrated original model in the literature. We emphasize the necessity to explore a range of components that influence the agents' decision making process and demonstrate that this is achievable through an evolutionary process if the rules are modularized as required. The end result is a set of plausible candidate solutions that closely fit the real-world data, which can then be nominated by domain experts.

CCS CONCEPTS

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

KEYWORDS

genetic programming, agent-based modeling, theory discovery, calibration, Artificial Anasazi

ACM Reference format:
Chathika Gunaratne and Ivan Garibay. 2017. Alternate Social Theory Discovery Using Genetic Programming:
Towards Better Understanding the Artificial Anasazi. In *Proceedings of GECCO '17, Berlin, Germany, July 15-19, 2017,* 8 pages.
https://doi.org/http://dx.doi.org/10.1145/3071178.3071332

GECCO '17, July 15-19, 2017, Berlin, Germany

ACM ISBN 978-1-4503-4920-8/17/07…$15.00
https://doi.org/http://dx.doi.org/10.1145/3071178.3071332

1 INTRODUCTION

Agent-based modeling has helped study emergent phenomena in a variety of complex systems including social science, ecology, biological systems, information science and economics. ABM is a bottom-up modeling technique, meaning macro-level phenomena are the emergent results of micro-level agent interactions. In addition to re-creating and predicting possible macro-behaviors of complex systems, ABMs can be used in reverse in order to understand which micro-behaviors are able to reproduce patterns in data observed in the real-world.

Typically, agents' micro-behaviors are defined using domain expert knowledge and/or assumptions made by the modeler. In many cases, however, multiple theories may exist, describing how the agents interact in real life; for example, the comparison of polarization theory, nearest neighbor theory and several formulated intermediate theories of fish schooling against data in [11]. To address this dilemma, a modeling protocol known as pattern-oriented modeling (POM) has been introduced [9, 25]. According to POM, modelers are required to construct multiple candidate models embodying existing theories of micro-behaviors. Simulations of these candidate models are then used to attempt to generate patterns from the simulated data that match patterns observed in the real-world data. Candidate models that match the real-world patterns the closest are then selected as winners, possibly explaining the phenomena being studied. POM is only one formalization of this approach. For example, in economics it is common to fine tune micro-behaviors of agents (such as trading rules) in order to match sets of established stylized facts, prior to using the models for experimentation [1, 7].

However, this methodology has a few disadvantages. Firstly, manual development of multiple candidate models can increase the risk in implementation and developer dependent discrepancies between candidate models [26]. Second, without proper modularization of the models, it can be non-trivial for modelers and domain experts to understand which elements of the winning models helped generate the preferable patterns. Such explanatory information can be useful for domain experts to verify and refine theories and possibly recombine multiple theories, evolving improved candidate models, which can better describe the real-world system.

In this paper we propose evolutionary social theory discovery, an automation of POM using using genetic programming to match patterns in real-world data. Through proper modularization of agent micro-behaviors, we apply genetic programming to evolve alternate micro-behaviors, automatically creating populations of candidate models as recommended in POM. When the objective of the genetic program is to fit patterns observed in real-world data, candidate models with higher possibility of representing the actual micro-behaviors of the real world are selected.

We demonstrate the utility of evolutionary social theory discovery by evolving the farm seeking strategies of agents in the Artificial Anasazi model of the Long House Valley [6]. The original model theorizes that the Anasazi only considered closeness of potential farm sites (sites with crop sufficient to support their household) to their current farm plot, when choosing their next farming location. We argue that there may have existed alternate theories describing the farm selection process and that these theories can be discovered through evolution towards the objective of matching archaeological records of annual population of the Anasazi in the Long House Valley. We evolved populations of candidate NetLogo models, essentially modeling alternate realities of the Anasazi with different farm-seeking strategies. As the evolutionary process progressed, the genetic program selected candidate models which better reproduced the archaeological records.

Our results demonstrate that there exist multiple decision trees which are able to match the data to substantially similar fitness. This indicates that alternate theories to the Anasazi farm selection behavior may exist, which must be further verified and nominated by domain experts (archaeologists in this example).

2 BACKGROUND

2.1 Related Methodological Work

Parameter calibration of ABMs has been explored in the literature through a variety of search algorithms. Full Factorial Design is perhaps the most straightforward approach. However, performing an exhaustive parameter search quickly becomes too computationally expensive on larger parameter sets, greater parameter ranges and with continuous variables. Alternatives have demonstrated to be much more efficient, such as full factorial design, latin hypercube sampling [14], and systematic design of experiments [15]. Further, genetic algorithms have been used to optimize ABM parameters to achieve a target macro-behavior for the population. Stonedahl used genetic algorithms to calibrate the parameter settings of a flocking model inspired by Reynolds boids model [19] to generate vee-formations [23]; Genetic algorithms have been used to calibrate the parameters of the Artificial Anasazi model to match archaeological population records [22] and calibrate parameters of ant foraging behavior emergent in the "Ants" model [4]. Finally, agent-based modeling software are now accompanied by calibration packages such as BehaviorSearch for NetLogo and OptQuest for AnyLogic.

Despite a considerable amount of literature on the calibration of ABMs using various search techniques, we focus on the evolution of the rules governing the agent behaviors themselves. Evolution of agent rules has been performed in the past through classifier systems, reinforcement learning [21], evolutionary neural networks [10] and co-evolution [18, 24]. However, recent studies have demonstrated the importance of evolving expressively defined micro-behavior rules to maintain sufficient evolvability within the population of candidate solutions [20]. Explicit and expressive representations also help ease model verification by domain experts in contrast to black-box learning techniques such ANN [27]. Expressive genetic programming has been used in evolving exit selection in crowd evacuation [28], human-environment action decision making [16], understanding bounded rationality in human decision making [17] and understanding human decision making in behavioral finance [5].

2.2 Artificial Anasazi

We illustrate our case using a well studied agent-based model from the literature, the Artificial Anasazi. The Kayenta Anasazi, a prehistoric race that occupied the Long House Valley in northeastern Arizona from 1800 BC to 1300, have been studied extensively as part of the Long House Valley Project conducted by the Museum of Northern Arizona and the Laboratory of Tree-Ring Research at the University of Arizona [6].

The Artificial Anasazi model consists of a spatial environment with seven zones: The General Valley Floor, North Valley Floor, Mid and North Valley Sand Dunes, Mid Valley Floor, Non-Arable Uplands, Arable Uplands, and Kinbiko Canyon. The primary agent in the model represents a typical Anasazi household of five persons. The household agents use harvests gathered from their farm plot to satisfy their demand for food. Farm plots can only consist of one farm and no households. The harvest is a stochastic product of the quality of the farm plot and the expected yield from the plot (this data is provided through archaeological records). Households can stock food surplus from the previous year. Every year, a decision is made on whether to find a new farm plot if the harvest from the current plot is insufficient to satisfy the demand of the household. Agents searching for a new farm plot try to farm on the closest possible patch to its old farm that is able to provide a harvest sufficient to support its household. Once the new farm is found, the household tries to find a suitable settlement by attempting to settle down as close as possible to a water source near the new farm plot. Unlike farms, multiple households can occupy the same patch. If no new suitable farm plot or settlement is found the agent will die.

The Artificial Anasazi model has a few important sub-models, out of which we select the farm selection sub-model for investigation. In the original model, it is assumed that the best farm plot is selected as the closest potential farm plot to an agent's current farm. In order for a patch to be considered as a potential farm by a household it must have no pre-existing farms or households and it should potentially produce enough yield to satisfy the household's minimal nutrition requirement.

Janssen [12] and Stonedahl & Wilensky [22] calibrate the parameters of the Artifical Anasazi model through minimization of the L^2 error between the simulated output of the number of households and the number of households reported through archaeological findings. The L^2 error is calculated as the Euclidean distance between the two time series data. The L^2 error is calculated as in equation 1.

$$L^2 error = \sum_{t=0}^{T} \sqrt{(n_t^{historical} - n_t^{simulated})^2} \quad (1)$$

Where n refers to the number of households, t refers to the current time step and T refers to the total number of time steps that the simulation is run for.

We choose the Artificial Anasazi model for the same reasons illustrated in [22]; unlike many ABMs encountered in the literature, the Artificial Anasazi is a well known, data-driven ABM, designed to fit a historical time-series dataset. Further, the model has gone through a thorough design process, following the Overview Design Details (ODD) protocol [8]. Finally, there have been documented attempts to calibrate the model, starting from Axtell et al. [2], to Janssen's factorial experiment design [12], to Stonedahl's genetic algorithm search [22]. Thus, we see our approach to evolve the rules of the artificial Anasazi model as a timely extension of the efforts to further understand the patterns in the archaeological records.

3 METHODOLOGY

The implementation of the Artifical Anasazi used for this study was obtained from the NetLogo 6.0 model libraries and is also available in the OpenABM model repository. Archaeological data files for Population counts, APDSI (Adjusted Palmer Drought Severity Index), and Agricultural productivity measures accompanied the model. ECJ, an evolutionary computation toolkit written in Java, was used to run the genetic programming (GP). Each resulting tree from the genetic program was then automatically written into NetLogo model files through NetLogo controlling API, to initialize and run the evolved NetLogo models.

Trees evolved from the GP setup represented grammars for NetLogo syntax and were translated into NetLogo syntax during the evaluation of each individual. The decision tree used for farm selection was modularized into GP nodes, consisting of functions and terminals. The function set used was as follows:

{ *MinOf, MaxOf, +, -* }

while the terminal set comprised of:

{ *CurrentFarm, PotentialFarmSet, CompareQuality, CompareDryness, CompareYield, CompareHouseholds, CompareHydro, CompareDistance*}

NetLogo uses a functional programming style syntax and is easily modularized into GP-node-ready functions. Table 1 maps each terminal and function node mentioned above into its corresponding NetLogo syntax. In order to generate legal syntax, the GP nodes were designed to be strongly typed; corresponding types are also listed in table 1. The comparator nodes encoded NetLogo statements that defined which attributes of potential farm plots were to be compared by the MinOf and MaxOf nodes. Comparator nodes defined comparisons of normalized values of the sensor inputs being compared so that they could be fairly added and subtracted from each other for trees that used multiple comparator types. The values *maxQuality*, *maxDry*, *maxYield*, *maxHouseholds*, *maxHydro* and *maxDistance* stand for the maximum values for their respective parameters for normalization.

The original Artificial Anasazi model uses a list data structure to hold the potential farm sites. This list is then iterated to find the closest potential farm site to the agent's current farm, which is selected as the new farm site. A new patch for the agent's household is then selected based on the selected new farm site and available water resources in the vicinity. Instead, in order to make use of NetLogo's inherent functions and to simplify the farm seeking behavior into a single line of code, we used a set representation of the potential farms to perform the same task. The following line of code allowed us to obtain a set of farms with the same criteria as

Table 1: Functions and Terminals used in evolutionary pattern-oriented modeling of farm selection of the Artificial Anasazi.

Node	Syntax	Return Type
MinOf	min-one-of (*patches*)[*comparator*]	*patch*
MaxOf	max-one-of (*patches*)[*comparator*]	*patch*
+	*comparator + comparator*	*comparator*
-	*comparator - comparator*	*comparator*
CurrentFarm	patch farm-x farm-y	*patch*
PotentialFarmSet		patches
CompareQuality	(quality / maxQuality)	*comparator*
CompareDryness	(APDSI / maxDry)	*comparator*
CompareYield	(yield / maxYield)	*comparator*
CompareHouseholds	((sum [num-occupying-households] of patches in-radius (water-source-distance * 2)) / maxHouseholds)	*comparator*
CompareHydro	((mean [hydro] of patches in-radius (water-source-distance * 2)) / maxHydro)	*comparator*
CompareDistance	((distance patch x-of-farm y-of-farm) / maxDistance)	*comparator*

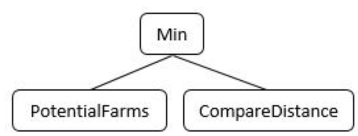

Figure 1: The original decision tree for the selection of the next best farm site for agents in the Artificial Anasazi model, using NetLogo's set representation.

the potential farm sites in the original model:

let potential-farm-set patches with [not (pxcor = farm-x and pycor = farm-y) and (zone != "Empty") and (num-occupying-farms = 0) and (num-occupying-households = 0) and (base-yield >= household-min-nutrition-need)]

Where, *farm-x* and *farm-y* are the coordinates of the agent's current farm. We could then use the *min-one-of* command in NetLogo to perform the same best potential farm site search as in the original model by finding the potential farm patch with the minimum distance to the current farm patch. A visualization of the decision tree used in the set representation of the original model is shown in figure 1 using the node representations in table 1.

It was assumed that all agents in the simulation had full information of the agricultural and social situation of the Valley. In previous studies, agents would search for the closest farm and possible water

Table 2: Translations of the five best decision trees into simplified NetLogo instructions.

Tree	Simplified NetLogo instruction
1	max-one-of (potential-farm-set) [- (yield / maxYield) - 3* (quality / maxQuality) - (APDSI / 9) - 2*((distance patch x-of-farm y-of-farm) / maxDistance)]
2	max-one-of potential-farm-set [((sum [num-occupying-households] of patches in-radius (water-source-distance * 2)) / maxHouseholds) + (APDSI / maxDry)]
3	max-one-of potential-farm-set [((distance patch x-of-farm y-of-farm) / maxDistance)]
4	max-one-of potential-farm-set [(APDSI / maxDry)]
5	max-one-of potential-farm-set [(yield / maxYield)]

source during re-settlement, and would continue to search until their requirement was met or would perish. By assuming full information, we allowed for this persistent search to remain unchanged.

The objective of each simulation was to minimize the L^2 error between the historical households count from archaeological data and the simulated household count over time. Accordingly, the fitness of each individual evolved from the GP was reported as the final L^2 error reported by the resulting NetLogo model after running for 551 years (from 800 AD to 1351 AD).

4 EXPERIMENTS AND RESULTS

4.1 Micro-behavior Discovery

We performed 5 genetic programming runs for the evolutionary micro-behavior discovery of the farm seeking behavior of the Artifical Anasazi agents. The GP was run for 50 generations. Each resulting GP individual was written into the NetLogo model and evaluated over five model runs. The individual's fitness was then reported as the average L^2 error of these five runs. We used the Half-and-Half tree builder by Koza [13] to initialize our GP population with a maximum depth of 8 and a minimum depth of 2, to prevent agents selecting their current farm again. The population size was set to 20 individuals. The GP was set to retry breeding 100 times if duplicates were generated in order to maintain diversity.

For model evaluations during the rule discovery process, we used the parameter settings which were obtained through Stonedahl's Calibration-1 experiment [22] (also in table 3), which provide the best individual fit to the data in the literature.

Figure 2 displays the Koza adjusted fitness of the best individual micro-behavior set that was discovered so far, by generation. It can be seen that four out of five runs converge quite quickly. It can be argued that this is due to the small population size used, leading the GP to discover a good solution from the modules that were available within the population at that time. Figure 3 displays the size of the best individual so far in each GP run. Four out of the five runs maintain quite small trees, in contrast to one of the runs which maintains a size of 23 nodes (figure 4). Table 2 provides translations of the five best decision trees into simplified NetLogo code.

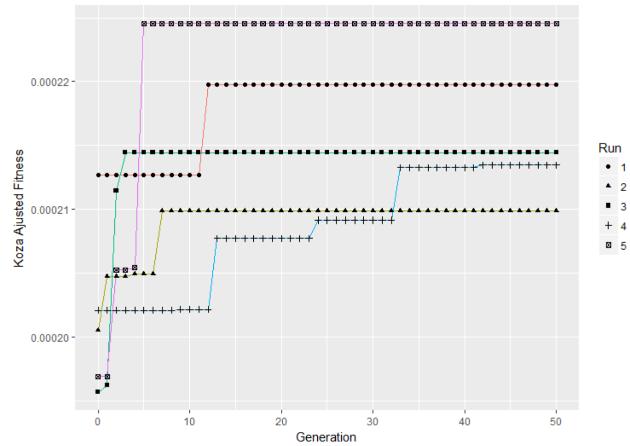

Figure 2: The Koza Adjusted fitness of the best GP individual so far in the evolution of farm seeking behavior of the Artificial Anasazi ABM.

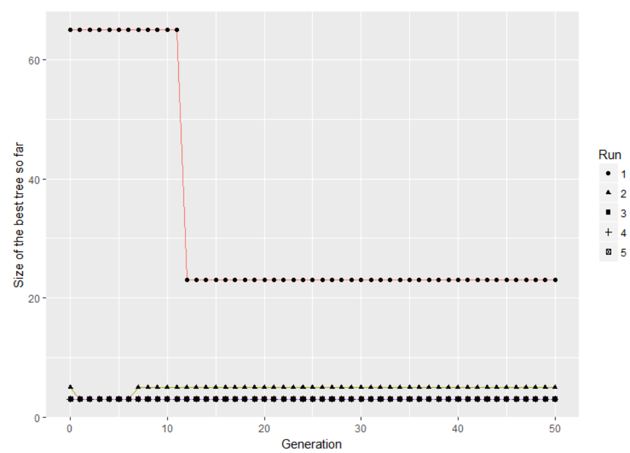

Figure 3: The size of the best GP individual so far in the evolution of farm seeking behavior of the Artificial Anasazi ABM.

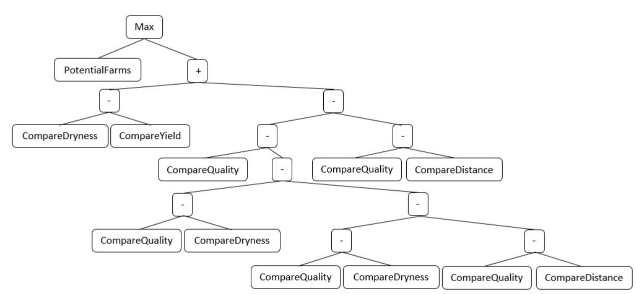

Figure 4: Tree 1 obtained through evolutionary POM on Artifical Anasazi farm seeking behavior. Phenotype: ***max-one-of (potential-farm-set) [- (yield / maxYield) - 3* (quality / maxQuality) - (Dryness / maxDryness) - 2*((distance patch x-of-farm y-of-farm) / maxDistance)].***

4.2 Calibration

We calibrated the parameters of the resulting five decision trees using the BehaviorSearch tool that is provided with NetLogo. BehaviorSearch utilizes a genetic algorithm (GA) to perform parameter calibration for NetLogo models. For each calibration, populations of 90 were used with mutation rates of 0.05 and crossover rates of 0.7. Tournament selection was used with tournament size of 3 and genes were coded using Gray coding. Calibrations were run for 200 generations. Similar to the Calibration-1 experiment in [22], we used only one evaluation of the model per genome; instead of assigning fitness as an average across multiple runs, in order to accommodate for the chaotic nature of ABM output [3] and to discover model outcomes (instead of outcome aggregates) that closely matched the historical data. 30 best replicate runs were performed every time the GA would discover a 'best so far' parameter set, in order to calculate the standard deviation of the model for the calibrated parameter set.

The parameters obtained through the calibration process are shown in table 3. Their respective replications, run 100 times each, are displayed in figure 5. The L^2 error statistics (the best L^2 error found by BehaviorSearch and the respective mean and standard deviation for the best parameter set over the 30 replicate runs) found through the calibration process for the five best decision trees are also shown in table 3. Additionally, Pearson product-moment correlation coefficients (r) were calculated for simulated time-series of each of the five candidate trees against that of the archaeological data, comparing the qualitative similarity of the shapes of the simulated time-series data against the archaeological data and are also found in table 3. Again, this method of evaluation was adopted from [22] for ease of comparison.

Our results indicated that Tree 3 had a best L^2 error (L^2 error = 715.097) less than that reported in previous studies. Further, the all candidate Trees demonstrated strong correlation against the archaeological time-series with Tree 2, Tree 3, Tree 4, and Tree 5 having stronger correlations than the best correlation reported in the literature.

4.3 Sensitivity Analysis

Sensitivity analysis can give valuable insight into the robustness of an ABM [26]. Both Janssen and Stonedahl have performed sensitivity analyses on the models resulting from their parameter searches. We use Stonedahl's approach for sensitivity analysis by using a GA to maximize the L^2 error reported by the simulation [22]. We used the same parameter ranges for the sensitivity analysis as used by Janssen and Stonedahl (10% variations). The results for the sensitivity analysis can be seen in figure 7, where the ±10% range is shown by the error bars, and the boxplots represent the range of parameters for which the worst fit was found. It is apparent that Tree 3 has a relatively high sensitivity to most of the parameters. Harvest adjustment shows high sensitivity for Trees 1, 3 and 5, while all models show moderate sensitivity to base nutrition need, fertility parameters and 'maize gift to child'.

5 DISCUSSION

Automating the theory discovery process provides us with some useful insights into the farm seeking behavior of the Anasazi agents. First, consider the micro-behaviors evolved from the GP tree on a modular basis. The modules used in this study include sensory information of the environment available to the agents; basically all of the nodes returning *comparators*. The *MinOf* and - operators indicate a negative correlation, and *MaxOf* and + indicate a positive correlation, of the sensory input to the preference towards a potential farm site. Therefore, by comparing the number of negative and positive correlations of each module in the resulting trees, we can make a weak inference on each sensor's impact on the preference towards a potential farm plot. Accordingly, when considering all five evolved trees, quality has the highest net negative correlation of three negatively correlated nodes. Dryness is positively correlated twice and negatively correlated once, indicating a slight positive correlation towards farm site preference. Yield remains neutral, while social presence (*CompareHouseholds*) is positively correlated towards farm site preference. Finally, distance to current farm, as assumed in the original model, exhibited a slight net negative correlation to farm preference (i.e.: closer potential farm sites are preferred). With more GP runs to support this evidence, this information can prove valuable towards building a comprehensive theory, which considers multiple sensory input, on how the Anasazi performed farm selection.

Next, consider the decision trees individually. A surprising result of the alternate theory discovery process was the emergence of Tree 3 which describes the counter-intuitive decision to seek for the furthest potential farm from the current one, as the next farm site. What is more surprising, is that BehaviorSearch reported a run with L^2 = 715.097 for this tree during calibration, which is a slightly better fit than the best run of the GA calibrated original model (L^2 = 733.6) in [22], despite exhibiting the complete opposite farm selection behavior. Tree 4 can also be considered counterintuitive , as the agents selected the driest patches to farm on. The emergence of such counter intuitive decisions to that used in the original model indicates a lack of selection pressure. In other words, the effect that distance had on the farm selection process, in terms of matching the historical population records, is less than anticipated. From a modeler's perspective, knowing the degree to which such counter intuitive theories, may fit patterns in the data, can provide insight into the robustness of the order in which modules defining the micro-behaviors are arranged. It may also indicate heterogeneity of the actual Anasazi farm selection strategy; the archaeological time-series data could have resulted from a forest of different micro-behaviors instead of a single strategy, reducing selection pressure on any particular tree of modules if considered homogeneously. Speciation for the inclusion of multiple farm seeking stratgies within the same Anasazi model, would help test this explanation and is a consideration for future studies. Evolving other sub-models of agent behavior, such as reproduction of new households and selecting new settlements, could also lead to the discovery of better fitting micro-behavior sets for the Artificial Anasazi.

Further, a closer look at variations between the calibrated parameter sets for each candidate model can provide interesting explanations for the agent behaviors. The best run of Tree 4 results from the lowest agricultural nutrition requirement among the models and the lowest minimum death age. This could be interpreted as follows, if the Anasazi were to prefer drier locations to farm and live

Table 3: Parameters resulting from the calibration of the five best trees that produce the lowest L^2 error and their best fitness, average fitness, and maximum Pearson's product-moment correlation coefficient to the archaeological time-series data, compared against the parameters for the run with the best fit of the original decision tree in the literature (Calibration-1 Experiment [22]).

Parameter	Tree 1	Tree 2	Tree 3	Tree 4	Tree 5	Original
HarvestAdjustment	0.53	0.51	0.65	0.52	0.62	0.64
HarvestVariance	0.45	0.46	0.43	0.43	0.43	0.44
BaseNutritionNeed	175	160	180	155	185	185
MinDeathAge	33	32	39	27	29	40
DeathAgeSpan	7	4	6	5	0	10
MinFertilityEndsAge	35	37	26	30	26	29
FertilityEndsAgeSpan	3	5	2	5	1	5
MinFertility	0.17	0.15	0.17	0.17	0.18	0.17
FertilitySpan	0.09	0.02	0	0.05	0.04	0.03
MaizeGiftToChild	0.35	0.34	0.28	0.2	0.28	0.47
WaterSourceDistance	17.0	2	23.5	12.0	15.0	11.5
Best L^2	784.242	837.557	715.097	791.319	803.046	733.6
Mean L^2 (σ)	1383.87(669.54)	1701.44(631.31)	1928.84(564.63)	1568.86(674.73)	1675.71(688.50)	965.42(275.9)
Max Correlation (r)	0.734	0.884	0.895	0.905	0.889	0.83

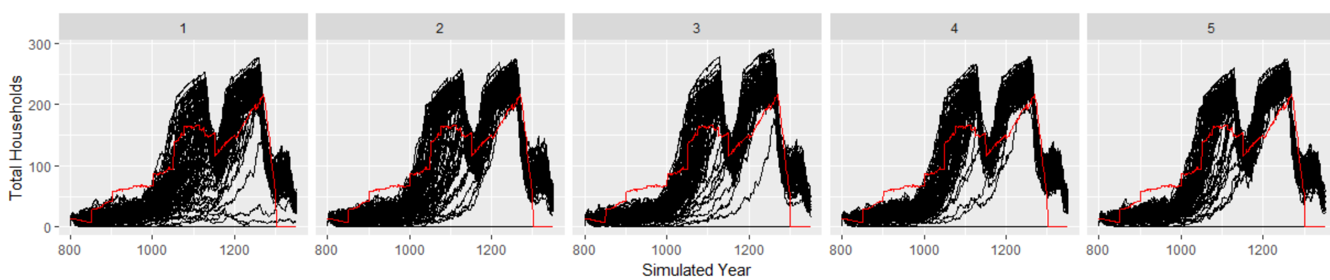

Figure 5: Comparison of Household counts over time for 100 runs of each best farm seeking behavior evolved through the GP against historical data (Red).

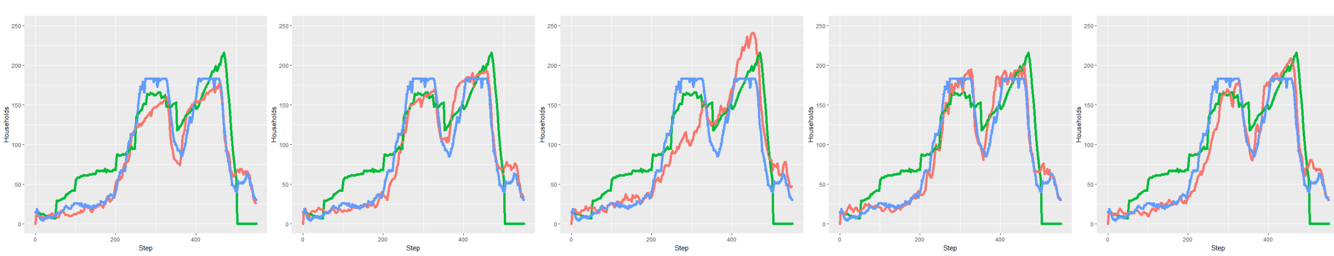

Figure 6: Comparison of Household counts over time for the best runs of each decision tree evolved from the GP (Red), against the historical data (Green) and the Stonedahl and Wilensky calibrated model (Blue).

in (settlements built close to farms), they would require a substantially lower nutrition need per household through agriculture and would probably have had shorter lifespans. Tree 2 has a much lower 'water source distance' parameter value than the other models. Tree 2 seeks to maximize the number of households around the new farm site, by estimating the households twice the preferred 'water source distance' away (households that may share the same water source). This can indicate the presence of tightly knit communities in comparison to large widespread communities. Social presence has not previously been considered in the farm seeking behavior of the original model.

Finally, qualitatively comparing the shapes of the time-series graphs in figure 6 with regard to the two peaks in population, reveals two plateaus formed by the original, GA calibrated decision tree, (which selects the closest potential site to farm in next), quite unlike the more triangular peaks of the archaeological data. In contrast,

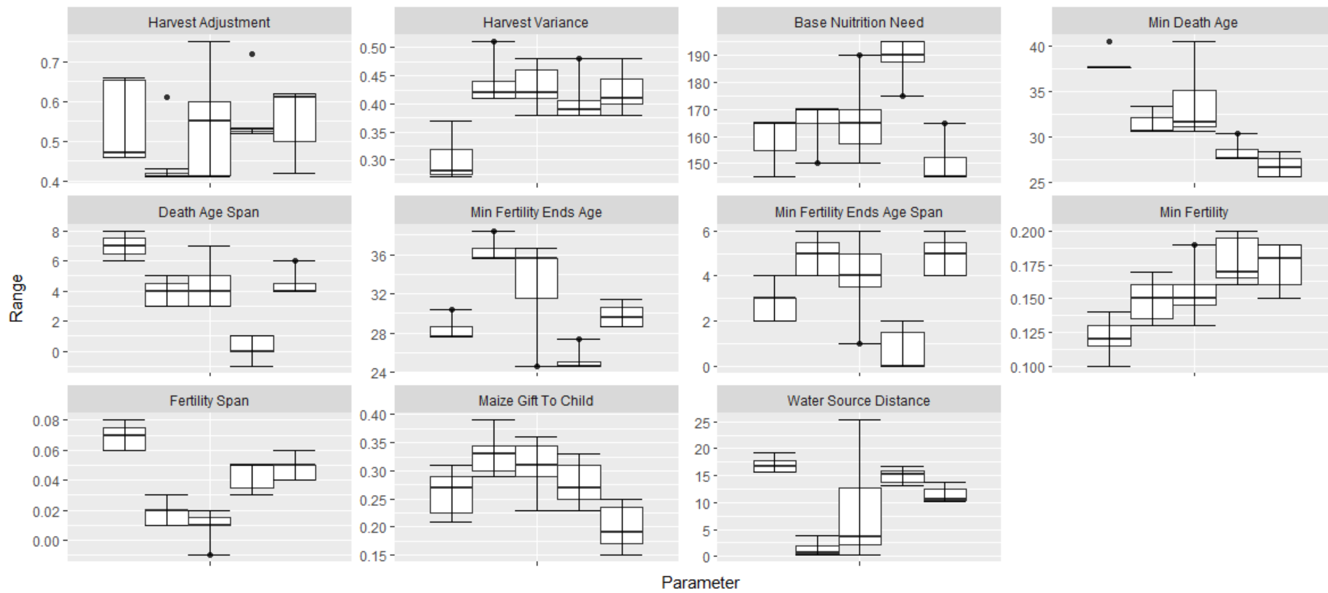

Figure 7: Results of sensitivity analysis on the five trees (ordered 1 to 5 from left to right) for each parameter. The error bars indicate the ±10% range used for sensitivity analysis and the boxplots show the range for which the fitness of the resultant model is at the worst possible.

the GP evolved trees show a similar shape to the peaks of the archaeological data. Tree 4, shows the closest match in shape when measured by the correlation coefficient ($r = 0.905$), inferring that the Anasazi may have preferred locations with highest yield, at least during the two peaks. In other words, preferring sites with better yield than closer ones helped break out of the 'population limit' observed in simulations of the original model. This was possibly caused by limited exploration of the valley by agents when using the original strategy, due to the requirement of always searching for closer, instead of searching for better, farming locations.

6 CONCLUSION

This paper presents an evolutionary approach to discovering alternate social theories for explanatory agent-based models, which are able to generate similar, and in some cases better, fit to real world data as the theories and assumptions upon which a model is original designed. We argue that this automated alternate theory discovery process is necessary to guide the model building process, avoiding blind assumptions and helping to inform domain experts on how well candidate theories may fit the data. The Artificial Anasazi model was used to illustrate this methodology in practice and demonstrate that multiple candidate models of farm selection may exist with similar or better fit to patterns in real-world data. We believe that we have shed light on the fact that the farm selection process may not have been monotonously controlled by proximity to an household's current farm, but may have been more complex, influenced also by other environmental factors such as social presence and potential yield.

Through this study, we bring to attention that modelers must be aware of the existence of alternate, plausible micro-behaviors. If model designers use an automated strategy to assess the importance of different sensory information available to an agent towards a decision making process, it will provide them with a dashboard of modules to construct candidate micro-behaviors along with their respective fitness and calibrated parameters. The resulting candidate models can then be nominated as theories by domain experts. Without testing other candidate theories, one may end up forcefully calibrating a model with rules that do not represent real world mechanisms.

In this study, we have not performed parameter calibration during the micro-behavior evolution, instead we calibrate the resulting models after the genetic program has discovered them. However, this may cause certain theories, which would have potentially provided a closer fit with the correct parameter set, to be siphoned out of the GP population during the theory discovery process. The simulations were run on a Dell Precision 3620 Mini Tower (6th Gen Intel Core i7-6700 3.40Ghz, 32Gb RAM), and it took approximately 2 CPU days to complete one experiment. Therefore, performing parameter calibration alongside the genetic program would have made this process impractical. A future direction for this work is attempting to integrate parameter calibration into the GP driven theory discovery process. Finally, we have only considered one pattern as our target data, the archaeological household estimates. Yet, a multi-objective search, targeting multiple patterns (such as the spatio-temporal distributions of households) would be more suitable and viewed as the next step for this study.

REFERENCES

[1] Robert Axtell, Robert Axelrod, Joshua M Epstein, and Michael D Cohen. 1996. Aligning simulation models: A case study and results. *Computational & Mathematical Organization Theory* 1, 2 (1996), 123–141.

[2] Robert L Axtell, Joshua M Epstein, Jeffrey S Dean, George J Gumerman, Alan C Swedlund, Jason Harburger, Shubha Chakravarty, Ross Hammond, Jon Parker, and Miles Parker. 2002. Population growth and collapse in a multiagent model of the Kayenta Anasazi in Long House Valley. *Proceedings of the National Academy of Sciences* 99, suppl 3 (2002), 7275–7279.
[3] Daniel G Brown, Scott Page, Rick Riolo, Moira Zellner, and William Rand. 2005. Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographical Information Science* 19, 2 (2005), 153–174.
[4] Benoît Calvez and Guillaume Hutzler. 2005. Automatic tuning of agent-based models using genetic algorithms. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*. Springer, 41–57.
[5] Shu-Heng Chen and Chia-Hsuan Yeh. 1999. Genetic programming in the agent-based modeling of stock markets. In *Proceedings of the Fifth International Conference on Computing in Economics and Finance, Boston College, MA, USA*.
[6] Jeffrey S Dean, George J Gumerman, Joshua M Epstein, Robert L Axtell, Alan C Swedlund, Miles T Parker, and Stephen McCarroll. 2000. Understanding Anasazi culture change through agent-based modeling. *Dynamics in human and primate societies: Agent-based modeling of social and spatial processes* (2000), 179–205.
[7] Domenico Delli Gatti, Corrado Di Guilmi, Edoardo Gaffeo, Gianfranco Giulioni, Mauro Gallegati, and Antonio Palestrini. 2005. A new approach to business fluctuations: heterogeneous interacting agents, scaling laws and financial fragility. *Journal of Economic behavior & organization* 56, 4 (2005), 489–512.
[8] Volker Grimm, Uta Berger, Donald L DeAngelis, J Gary Polhill, Jarl Giske, and Steven F Railsback. 2010. The ODD protocol: a review and first update. *Ecological modelling* 221, 23 (2010), 2760–2768.
[9] Volker Grimm, Eloy Revilla, Uta Berger, Florian Jeltsch, Wolf M Mooij, Steven F Railsback, Hans-Hermann Thulke, Jacob Weiner, Thorsten Wiegand, and Donald L DeAngelis. 2005. Pattern-oriented modeling of agent-based complex systems: lessons from ecology. *science* 310, 5750 (2005), 987–991.
[10] Matthew Hausknecht, Piyush Khandelwal, Risto Miikkulainen, and Peter Stone. 2012. HyperNEAT-GGP: A HyperNEAT-based Atari general game player. In *Proceedings of the 14th annual conference on Genetic and evolutionary computation*. ACM, 217–224.
[11] Andreas Huth and Christian Wissel. 1994. The simulation of fish schools in comparison with experimental data. *Ecological modelling* 75 (1994), 135–146.
[12] Marco A Janssen. 2009. Understanding artificial anasazi. *Journal of Artificial Societies and Social Simulation* 12, 4 (2009), 13.
[13] John R Koza. 1992. *Genetic programming: on the programming of computers by means of natural selection*. Vol. 1. MIT press.
[14] Ju-Sung Lee, Tatiana Filatova, Arika Ligmann-Zielinska, Behrooz Hassani-Mahmooei, Forrest Stonedahl, Iris Lorscheid, Alexey Voinov, J Gary Polhill, Zhanli Sun, and Dawn C Parker. 2015. The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation* 18, 4 (2015), 4.
[15] Iris Lorscheid, Bernd-Oliver Heine, and Matthias Meyer. 2012. Opening the 'black box' of simulations: increased transparency and effective communication through the systematic design of experiments. *Computational and Mathematical Organization Theory* 18, 1 (2012), 22–62. https://doi.org/10.1007/s10588-011-9097-3
[16] Steven M Manson. 2005. Agent-based modeling and genetic programming for modeling land change in the Southern Yucatan Peninsular Region of Mexico. *Agriculture, ecosystems & environment* 111, 1 (2005), 47–62.
[17] Steven M Manson. 2006. Bounded rationality in agent-based models: experiments with evolutionary programs. *International Journal of Geographical Information Science* 20, 9 (2006), 991–1012.
[18] Matjaž Perc and Attila Szolnoki. 2010. Coevolutionary gamesâĂŤa mini review. *BioSystems* 99, 2 (2010), 109–125.
[19] Craig W Reynolds. 1987. Flocks, herds and schools: A distributed behavioral model. *ACM SIGGRAPH computer graphics* 21, 4 (1987), 25–34.
[20] Lee Spector and Nicholas Freitag McPhee. 2016. Expressive Genetic Programming: Concepts and Applications. In *Proceedings of the 2016 on Genetic and Evolutionary Computation Conference Companion*. ACM, 589–608.
[21] Peter Stone and Manuela Veloso. 2000. Multiagent systems: A survey from a machine learning perspective. *Autonomous Robots* 8, 3 (2000), 345–383.
[22] Forrest Stonedahl and Uri Wilensky. 2010. Evolutionary Robustness Checking in the Artificial Anasazi Model.. In *AAAI Fall Symposium: Complex Adaptive Systems*. 120–129.
[23] Forrest Stonedahl and Uri Wilensky. 2010. Finding forms of flocking: Evolutionary search in abm parameter-spaces. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*. Springer, 61–75.
[24] R Paul Wiegand, William C Liles, and Kenneth A De Jong. 2002. Analyzing cooperative coevolution with evolutionary game theory. In *Evolutionary Computation, 2002. CEC'02. Proceedings of the 2002 Congress on*, Vol. 2. IEEE, 1600–1605.
[25] Thorsten Wiegand, Florian Jeltsch, Ilkka Hanski, and Volker Grimm. 2003. Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and application. *Oikos* 100, 2 (2003), 209–222.
[26] Uri Wilensky and William Rand. 2007. Making models match: Replicating an agent-based model. *Journal of Artificial Societies and Social Simulation* 10, 4 (2007), 2.
[27] Ning Wu and Elisabete A Silva. 2010. Artificial intelligence solutions for urban land dynamics: a review. *CPL bibliography* 24, 3 (2010), 246–265.
[28] Jinghui Zhong, Linbo Luo, Wentong Cai, and Michael Lees. 2014. Automatic rule identification for agent-based crowd models through gene expression programming. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1125–1132.