The Data Absorption Technique: Coevolution to Automate the Accurate Representation of Social Structure

Deborah Duong

Agent Based Learning Systems
Birmingham, Alabama, USA
dduong@agentBasedLearningSystems.com

ABSTRACT

Coevolution is a technique that not only enables simulations to represent the formation of social structure, but also facilitates their capability to emulate the processes of iterative feedback underlying social data. With coevolution, simulations can be validated at a general level and automatically applied to specific scenarios, recreating the dissipative structure of real world social institutions from data alone. Examples of "seeding" coevolutionary systems with data are given, and techniques for deriving the causal processes of a system from a similar system are presented. Experimental results illustrate the significant advantage of using data absorption techniques to emulate the system that underlies correlative data.

Keywords: Cognitive agent based simulation, coevolution, data absorption, social science simulation, Bayesian optimization algorithm, irregular warfare.

1 INTRODUCTION

This paper presents a technique for mimicking a social system with a simulation when that system is not fully specified by the program and the data, called data absorption. It is a technique for filling in what is not there, not through extrapolation or other statistically based techniques, but by what makes sense to the forming system, according to the motivations of individual agents. Because it goes through motivation, the data absorption technique computes from first principles, explaining through process as opposed to describing statistically. Scientifically valid analysis requires explanation: in order to test policy on a social system you first must represent the causal processes of a system, and only once you have the basic causal processes correct can you test new policies that the system has not been exposed to. Statistical description and extrapolation cannot test the effects of new policies because it does not effectively embody causation. In data absorption, the motivations that caused the correlations in the data are captured, deriving cause from correlation.

The way that computational social science typically addresses causation is with a complete model of causal process: input data is fed through these processes, rather like a physics-based model. However, there is a fundamental difference between computational social science and physics: in physics we always know the process ahead of time, because of universal law. However, in social science, although general laws may be applicable, the exact processes and motivations behind them may change from scenario to scenario. Therefore a new system must be tailored to every individual case. However, if this is done by hand, then we no longer have a way to validate the simulation before the simulation is run on a particular scenario. Data Absorption facilitates validation by simulating only general social science concepts and allowing the data to fill in the details of the process. Validation can occur by tests that separate the testing set and the training set with no human in the loop to invalidate the technique. Automating the representation of details of process is necessary before simulation can be applied to the bigdata of the internet in a scalable manner.

Additionally, data absorption helps to ensure that the data is relevant and applicable to the processes in social simulations. There is a gap in computational social science simulations between theory-centric simulations, that find emergent patterns computed from lower level rules; and data-centric simulations, that represent the details of data in real world scenarios. Theory-centric simulations capable of emergence show proof of principle of general concepts, while the data-centric simulations tend not to form emergent patterns. Policy testing, however, needs both the theory-centric simulation's computation from assumptions for valid analysis, and the data-centric simulation's close representations to an existing scenario. The reason for the gap is the recognition that dissipative structure emerges from the processes of the simulation, and that the validity of the pursuit of computational social science depends on the lack of obviousness that the upper level emergent pattern would form from the lower level rules, lower level rules which in combination should be nonlinear and more than the sum of their parts. When a scientist simulates from first principles, he loses control of the outcome, and that is how it should be for valid analysis, because to control the outcome is to put the answer to the question in the question, to “rig” and to lose objectivity. Social simulation is simulation of possible worlds, but to do policy analysis the scientist must emerge the details of this particular world, using some technique other than running simulations until the details of a real world scenario emerge by sheer coincidence. Data absorption can help find the lower level patterns that caused the data so that they emerge by the same processes in the real world to be simulated.

2 COEVOLUTION

To know how to mimic a system with incomplete data we look to complex adaptive systems in nature that mimic other systems. One place to look is the way a small child learns language. Cognitive scientists note that there is very little explicit correction of the child’s language. Somehow the models in the child’s head comes to mimic his parents and teachers, without full explication of the underlying model. We are trying to make social simulation do the same thing: using an incomplete set of correlative data fill in the processes that make sense.

Some natural language processing (nlp) methodologies mimic this process. One is the author’s nlp program, Indra, from which ontologies emerge from parsed freetext. Indra uses a coevolutionary technique to fill in an ontology when only partial information is known. Ontologies emerge from Indra through a process of “seeding.” Indra uses actor-action-object triplets in text to categorize the actors according to the actions they perform, simultaneously categorizing actions according to the actors that perform them. The categories formed are then treated as new actors and actions. Because of the feedback between the categorization of actors and the categorization of actors, only a few exemplars in the categories are needed to form an ontology of actors and actions, an ontology which can be used to find new individuals that belong to a category. If we “seed” a developing Indra ontology with examples of individuals, such as Osama bin Laden and Sadaam Hussein, Indra will find actions these individuals have in common, use these actions to find other individuals of the same type, and then use the new set of individuals to expand upon the common actions, in iterative feedback [1].

Another place in nature to look to find the mimicking of the processes of a system is in examples of coevolution in natural ecosystems. Coevolution works similarly to Indra’s emulation of language learning, in that there is an iterative feedback. For example, if in the real world we were to place the predators of one ecosystem into another, the prey of the ecosystem would have to adapt to the new predator, say to run faster. Then of course, the predators already existing in the ecosystem would have to adapt to their prey by running faster as well. In this sense, the ecosystem is “seeded” with a species from another system, which causes the ecosystem to take on many more traits of the ecosystem that the species is from, than the species itself has. An ecosystem may be thought of a house of mirrors, in that the individual species contain the whole as well.

Coevolutionary simulations mimic seeding in natural ecosystems. For example, the author's Symbolic Interactionist Simulation of Trade and Emergent Roles (SISTER) uses coevolutionary pressure to impart culture to its members. Each agent in SISTER has its own autonomous genetic algorithm, that tells the agent its plans to trade goods, as well as a tag to display that communicates the agent's desired trades to the other agents. The tag comes to have meaning in terms of trade behavior by this process: Once an agent develops a corresponding trade plan with another agent and finds the agent by seeking the tag, as long as the trade is good for both agents, they will attempt to make the trade again. However, when they repeat the trade, they seek the tag again rather than the individual agent. The tag allows other agents to engage in the same trade. Because trade is good for agents, the agent seeking the tag puts selective pressure on all the agents to have a corresponding tag, a corresponding trade plan, and the goods to trade. Other agents respond to this pressure, get in on the trade, and start wearing the tag to indicate that they want to trade [2].

A tag can also mean the recipe for how to make goods: for example, if an agent combines several goods together, resells the composite good, and displays a tag to sell it; then all the agents which want to buy the composite good and all the agents that want to sell the parts that make up the composite good will learn the tag. Then, if a new agent is introduced in the simulation and it displays the tag, the new agent will receive coevolutionary selective pressure from the sellers of the parts of the composite good and the buyers of the good to make the composite good correctly. In SISTER, the system of trade is recreated in the individual new members, and even survives the deaths of all the individual members as long as there are as many births and they do not die at the same time [3].

This same coevolutionary pressure may be used to help the system of a data-centric simulation come to take on the traits of the real world system that made the data. SISTER is a theory-centric simulation as opposed to the data-centric simulations needed for policy analysis, but the same principles still apply. Nexus is a data-centric simulation that uses data adaptation to mimic the vicious and virtuous cycles of behavior that cause the real world data. However, instead of single agents learning an existing system as in SISTER, all Nexus agents receive coevolutionary pressure to learn the system at once.

Nexus is coevolutionary in that agents together develop corresponding plans, each of them having their own evolutionary algorithm with which to learn behaviors that increase their individual well being. Any set of corresponding behaviors that is expected by agents is a social institution, and in the presented example the social institution under study is corruption. Corruption is a social vicious cycle that people do not want to be involved with, but feel they must in order to participate in society: it is social because it involves expectations of other's behaviors, for example, one would not offer a bribe if they did not expect the bribe to be accepted.

Nexus Algorithm

Nexus incorporates a co-evolutionary genetic algorithm to model a dynamic role-based network. Nexus has been part of several important studies of Irregular Warfare. In the OSD Africa Study, Nexus represented corruption, and in the US Army TRAC Tactical Wargame (TWG), Nexus represented dynamic role networks of key leaders and terrorists from which emanated intelligence messages [4],[5],[6].

Every Nexus agent has a Bayesian Optimization Algorithm (BOA) that agents utilize to determine behavioral strategies. In the Africa Study, strategies included bribing and stealing behaviors. Agents also use the BOA and preferred attributes to determine network partners.

A Nexus iteration starts with each agent choosing network partners from a list of qualified candidates. In the Africa Scenario,analysts modeled 65 roles in three networks: kin, bureaucratic, and trade networks. For example, a young single male may qualify as a husband and choose a wife according to preferred attributes, such as gender, age, ethnicity, etc. Relationship attrition may occur at every cycle. Networks, behaviors, roles, and attributes are inputs to a Nexus scenario. Analysts determine whether attributes are subject to learning or not.

Agents behave according to role relations, and this behavior may be witnessed and revealed to other agents depending on role relations. For example, in the TWG scenario of Afghanistan, a terrorist might drink tea with his cousin, performing a behavior that may be observed by a store owner. The store owner may share this information with a human intelligence (HUMINT) agent, leading to the inference that the men are cousins.

In the Africa Scenario, a father might buy food for his child and give a bribe at the same time because of a food shortage. In Nexus, agents conserve money. Money flows through accounts, and corrupt actions are defined as inappropriate transfers from one account to another. Inappropriate transfers include action such as bribing an employer for hiring through a kick back, or police stealing money from citizens at a checkpoint.

Results from bribes and stealing are implemented in rules rather than hard coded. For instance, bribe results may differ depending on whether the bribe was witnessed or not, and whether the witness was friendly to briber or not. Agents gain utility as a result of behaviors that involve money transfers. The utility occurs upon direct consumption of a desired commodity. For example, in the Africa Scenario, utility occurs when the maternal grandmother of a matri-local tribesman eats fruit she bought at the market. As scenario input, analysts determine behaviors and roles that generate utility depending on the culture. Utility for the time that a strategy is in effect is the “fitness function”, the equation by which a strategy is evaluated, for the BOA.

The BOA tests 20 chromosomes (evaluation factors) regarding bribing, stealing, accepting bribes, accepting or rejecting network partners who bribe and steal every 20 days. An example of a chromosome is the number of times bribes were accepted or declined. Starting conditions for behavior frequency are based on demographics of the targeted real-world location. For instance, relatively corrupt societies would initially engage in greater amounts of bribing and stealing than relatively non-corrupt societies. Utility over the 20-day test is based on the amount of utility agents incurred during their role transactions. The utility is the fitness of the chromosome. Nexus agents test a single strategy for 20 days (a parameter of the simulation), and then rate the strategy in terms of utility and accomplished goals. The agent will then test a second strategy for the next 20 days, and then rate that strategy. This pattern of testing and rating strategies continues for 20 different strategies (another parameter of the simulation), each tested for 20 days, culminating after 20 x 20 = 400 days.

The BOA captures the 10 best strategies for future use and discards the 10 worst strategies. The number of strategies for future use and the number of strategies discarded are simulation parameters set by the analyst. The BOA utilizes the 10 best strategies to reformulate 10 additional strategies for further testing by the agent. The result is 20 randomly mixed strategies being tested by the agent starting on day 401 that include the 10 best strategies from days 1-400 and 10 additional strategies that are statistically similar to the 10 best strategies.

Representing behaviors as a function of role relations captures the processes of sociology. Nexus models change in role relations and behaviors well because it bases role changes on utility. Coevolving BOAs seek individual utility, forming social structures. For example, bribing is social because in order to offer a bribe, you must believe that the bribe will be accepted, or at least that the person bribed will not inform authorities. Therefore, agents must separately develop corresponding plans to bribe and expect bribes as part of the social environment. Modeling an emergent social institution such as bribing is akin to Adam Smith's invisible hand concept where markets self-correct to accommodate consumers. Nexus models IW concepts well because Nexus allows analysts to test courses of action against a natural system. For instance, an analyst can test whether interventions such as transparency programs replete with associated penalties influence corrupt practices or not. Additionally, natural forces at work such as consumer preference prevent stores from stealing from customers, else the customers will not return.

Nexus utilizes Bayesian networks to describe measurable phenomenon based on demographic attributes from the country of interest, such as the chance of bribing for a particular tribe in a particular part of a country. The Bayesian network captures relationships between agent attributes and behaviors in the real world.

The BOA that each agent utilizes is unique such that the BOA can proceed from any point in reality, and evolves from that point, unlike most evolutionary algorithms. Because the BOA is coevolutionary, agents expect each other to behave in accordance to their roles that are input from real-world data, and apply selective pressure on each other according to their expectations. For example, if agents expect Mongos from Congo to bribe, then they offer bribes to this class of agents, placing selective pressure on Mongos to maintain the expectation. However, the agents within a class generally converge on one strategy or another according to their utility. For example, if agents believed that 60% of Mongos from Congo accept bribes and 40 % do not, then Nexus starts off with 100 % of Mongo agents accepting bribes 60% of the time and not accepting bribes 40% of the time. However, as the simulation reaches equilibrium after six years, 60% of Mongo agents accept bribes 100% of the time, and 40% of Mongo agents accept bribes 100% of the time. They feed their utility by maintaining public expectations, and also by developing behaviors that are individually rewarded by their relations.

3 Data Absorption in Nexus

To absorb data, every agent in Nexus is initialized with tendencies to behave in a way that real world people do, given their demographic characteristics. However the agents are not required to behave this way, rather they are just "hints" at how to behave. They are free to evolve any path that benefits them, however they do experiment with the behaviors that exist in a real world system to test if they beneficial. It is not a sure-fire thing that the behaviors of their demographically similar counterparts in the real world will benefit them: they still have to develop similar networked relationships as their counterparts for the behaviors to benefit them. However, just experimenting with the same behavior as their counterparts make it more likely that they will develop those relationships. For example, a government service provider agent experimenting in accepting bribes may be more likely to provide services to a citizen agent that offered bribes, if the bribe was in both their benefit. It would be in the citizen's benefit, for example, if there was a shortage of the service so that the bribe "corrects the price" of the service. In Nexus, runs with shortages result in an increase of bribing behavior, and other processes in the complex system could cause a shortage endogenously. Or it could be in the citizen's benefit to bribe if all service providers demanded a bribe. Rather than force behavior, we give agents options of behavior, and let them choose based on utility. To maintain realism, every agent is given the distribution of behaviors expected for its demographic group and initially tries all of them in proportion. For example, if female members of the Mongo tribe accept bribes 60 percent of the time, then every female of the Mongo tribe starts out experimenting with accepting bribes 60 percent of the time and not accepting them 40 percent of the time. Because everyone is offering everyone else offers to bribe at the rate expected by the demographic characteristics of the population, they experience the same motivational space of the original system, and all they have to do is respond to incentives through the networks that they are able to develop, given the resources available to them.

This research indicates that when the outer behaviors of a system are fed to the system, they pick up the inner choices that cause them. The outer behaviors provide gradient for the corresponding plans in the society to form, while at the same time, they only form if they are actually the best choice for the individuals when they are tested. As the networks of the agents converge, individual agents typically settle on only one pattern of behavior. If the other facts about the simulation limit the resources so that only some may succeed in being part of the network of utility increasing corresponding behaviors, then only a subset of individuals of a demographic groups come to have the behaviors. In the above example, toward the end of the simulation, resource constraints would cause 60 percent of female mongos to bribe 100 percent of the time with their established connections, while 40 percent of female mongos would not be able to develop the connects to offer bribes. The research indicates that the data absorption technique facilitates the system reaching a steady state similar to the system that it emulates, so that the proportion of agents participating in behaviors expected of demographic groups remains constant throughout the simulation, and is consistent across simulations for the demographic groups as well. Importantly, the motivations of agents towards the end of the simulation in the network they develop with available resources causes the proportions, so that the emulated system is ready for the testing of policy when behaviors have converged to the steady state. Figure 1 illustrates the steady state of the agents of a simulation experiment that mimics the real world system through data absorption.

Gender	**Tribe**	**Location**	**Age**	**Sector**	**Residence**	**Education**	**Behaviors developed in R1**	**Behaviors developed in R10**
Female	Azande	Region4	Under15	Industry	Matrilocal	Income2	bribeforservices	bribe for services
Female	Azande	Region4	WorkingAge	Government	Matrilocal	Income6	accept bribe for services	accept bribe for services
Female	Azande	Region4	WorkingAge	Industry	Matrilocal	Income4	none	none
Female	Foreign	Region4	Under15	Government	Patrilocal	Income7	bribe for services and accept bribe for services	bribe for services and accept bribe for services
Female	Kongo	Region1	WorkingAge	Industry	Patrilocal	Income3	bribe for services	bribe for services
Female	Luba	Region3	WorkingAge	Industry	Matrilocal	Income3	none	bribe for services
Female	Luba	Region3	WorkingAge	Industry	Patrilocal	Income4	none	bribe for services
Female	Mongo	Region2	WorkingAge	Industry	Patrilocal	Income3	bribe employer and bribe for services	bribe employer
Female	Other	Region1	Under15	Government	Matrilocal	Income7	bribe for services	bribe for services
Female	Other	Region1	Under15	Industry	Matrilocal	Income4	bribe for services	bribe for services
Female	Other	Region1	WorkingAge	Industry	Matrilocal	Income2	bribe for services	bribe for services
Female	Other	Region3	Under15	Industry	Patrilocal	Income5	bribe for services	bribe for services
Female	Other	Region3	WorkingAge	Government	Patrilocal	Income9	none	bribe for services
Female	Other	Region4	WorkingAge	Government	Patrilocal	Income5	bribe for	bribe for

							services and accept bribe for services	services and accept bribe for services
Female	Other	Region4	WorkingAge	Industry	Matrilocal	Income4	bribe for services	bribe for services
Male	Azande	Region2	Under15	Government	Matrilocal	Income5	bribe for services and accept bribe for services	bribe for services and accept bribe for services
Male	Azande	Region4	Under15	Industry	Matrilocal	Income2	bribe for services	bribe for services
Male	Azande	Region4	WorkingAge	Industry	Matrilocal	Income3	bribe for services	bribe for services
Male	Foreign	Region3	WorkingAge	Industry	Neolocal	Income2	bribe for services	bribe for services
Male	Foreign	Region4	WorkingAge	Industry	Neolocal	Income3	bribe for services	bribe for services
Male	Luba	Region3	WorkingAge	Government	Patrilocal	Income6	bribe for services and accept bribe for services	bribe for services and accept bribe for services
Male	Luba	Region3	WorkingAge	Industry	Patrilocal	Income3	bribe for services	bribe for services
Male	Mongo	Region2	Under15	Industry	Patrilocal	Income3	bribe for services	bribe for services
Male	Other	Region1	Under15	Industry	Matrilocal	Income2	bribe for services	bribe for services
Male	Other	Region1	WorkingAge	Industry	Patrilocal	Income3	bribe for services	accept bribe employer
Male	Other	Region2	Under15	Government	Patrilocal	Income7	bribe for services	bribe for services
Male	Other	Region2	WorkingAge	Government	Patrilocal	Income6	bribe employer, accept bribe employer, bribe for services and accept bribe for services	bribe for services accept bribe for services
Male	Other	Region2	WorkingAge	Industry	Patrilocal	Income4	bribe for services	bribe for services
Male	Other	Region3	WorkingAge	Industry	Matrilocal	Income2	none	bribe for services

Figure 1. Individual agent demographic characteristics kept the same between two runs, while behaviors are learned. Two runs, R1 and R10, using the data absorption technique illustrate remarkable consistency in the development of corresponding behaviors across runs. This consistency is behind the statistic that the data absorption technique using the output simulation data is more consistent in deriving the real world simulation state then the data directly from the real world itself.

4 NEXUS EXPERIMENT

In this experiment, agents are initialized with the desired behaviors of the real world system to be simulated, in proportion to their demography, before they have a chance to develop the networks that support the behaviors. Two thirds of the runs gradually develop networks that support the behaviors through the motivations of the agents, and one third never do. It is hypothesized that the motivations captured do not tell the whole story, and so stochastically the steady state of the real world system was out of reach one third of the time. Note that this behavior is not consistent enough to test a policy change by holding all else the same. However, we can take the output from one of group of two thirds of simulation results that do in fact emulate the real world processes, and feed that back as input to another run of Nexus. This data is still partial, outwardly measurable data, as opposed to a check point restart type of process, in order to give the stochastic agents the option to develop other outcomes and in order to more closely cover the space of possible outcomes than is possible with checkpoint restart. When the new simulation runs are seeded with the output behaviors of the old, the desired real world processes develop significantly more consistently, enabling policy analysis that holds all else the same.

Figure 2 shows that the behavior from the data absorption run using output from another simulation is more consistent than the run directly using data directly from the country. Figure 3 and figure 4 show the differences between tests of policy on both the runs using country data and the runs using the data output from the country data simulations. These runs illustrate how not “keeping all else the same” by using the data absorption technique can skew the results of a test of policy, leading analysts to erroneous conclusions. In the example, the test of policy is the application of stiffer penalties for corruption and then removing that penalty three years after introducing it. In the runs with data straight from the real world country (figure 3), the policy is a failure. Without the penalty, the system seems to naturally decrease in corruption about a third of the time, and with the penalty, it only decreases in corruption about sixth of the time. However, the test was performed with the assumption that the simulated country started out with the dissipative structures of corruption in the real world, which was not the case in the runs using data from the country without the data absorption technique. The test is more accurately performed when the simulation is initialized to the dissipative structures in the real world through the data absorption technique.

In the data absorption runs (figure 4), the penalty resulted in fewer service bribes even after the removal of the penalty, leading the system to a new steady state, one fourth of the time. If the penalty was never applied, the society would improve on its own a tenth of a time. This indicates that the policy of stiffer penalties for corruption was somewhat of a success in a corrupt country. This comparison shows the improvement of the accuracy and validity of policy test analysis with the data absorption technique. However, note that the result does not advocate government intervention in the cases where the country does not start out with a significant level of corruption: according to the research results, in these cases, the government intervention could cause the country to actually become corrupt, perhaps by suppressing pre-existing natural limits to corruption.

Service Bribes			
	Country Data	Data From Run	
Like Original, Stayed Higher	19	20	
Unlike Original, Decreased	10	2	
		chi squared test	0.007832

Figure 2. A comparison of 29 runs directly from country demographic data with 22 runs come from the output of the run of one of the 19 simulations that, like the original real world data, retained the dissipative structures of the social institution of bribes for government services. This social institution was maintained with consistency across runs, at the p=0.007 level, and was maintained by virtue of the motivations of the agents toward their personal utility.

Service Bribes			
	without treatment	with treatment	
Stayed High	19	17	
Decreased	10	3	
		chi squared test	0.0870244

Figure 3. Service behavior under penalty applied then removed in country data runs. At the p= 0.08 level, which may in some cases be interpreted as a significant result, the policy appears to be a failure. However, the run was not consistently initialized with the dissipative structures of corruption that existed in the country. The data absorption technique is needed to do that.

Service Bribes			
	without treatment	with treatment	
Stayed High	20	21	
Decreased	2	7	
		chi squared test	0.0145238

Figure 4. Service behavior under penalty applied then removed in data absorption runs. At the p= 0.01 level, a strongly significant result, the policy was a success given that the country started out in a corrupt state, by virtue of the data absorption technique.

5 SUMMARY

Data absorption can increase the validity of a scientific analysis of policy by helping emulate the system under study through the particular social institutions that caused the correlative data ingested by the simulation. This technique can bridge the gap between data-centric and theory-centric simulations, facilitating emergence in simulations of particular real world scenarios. Its automation gives simulation the a greater potential to be applied in a relevant manner to the bigdata of the internet, by virtue of the fact that the simulation need not be individually tailored to different real world scenarios, and can undergo validation testing by separation of the testing set and the training set. This technique makes use of coevolution to seed a system with observable outward data, so that the inner motivations of agents are found and filled in to complete the model of social process.

.

REFERENCES

[1] Duong, Deborah, Nicholas Stone, Ben Goertzel, and Jim Venuto. 2010. "Indra: Emergent Ontologies from Text for Feeding Data to Simulations" *Spring Simulation Interoperability Workshop*. DOI = http://www.scs.gmu.edu/~dduong/SIWindra.pdf .

[2] Duong, Deborah Vakas and John Grefenstette. 2005. "SISTER: A Symbolic Interactionist Simulation of Trade and Emergent Roles". *Journal of Artificial Societies and Social Simulation*, Januaary. DOI= http://jasss.soc.surrey.ac.uk/8/1/1.html.

[3] Duong, Deborah Vakas, 2008. "The Generative Power of Signs: Tags for Cultural Reproduction" *Handbook of Research on Agent-Based Societies: Social and Cultural Interactions*, Goran Trajkovsky and Samuel Collins, eds. DOI=http://www.scs.gmu.edu/~dduong/GenerativePowerOfSigns.pdf.

[4] Duong,Deborah Vakas, Robert Turner and Karl Selke. 2010."Crime and Corruption," *ESTIMATING IMPACT: A Handbook of Computational Methods and Models for Anticipating Economic, Social, Political and Security Effects in International Intervention,* Kott and Citrenbaum, eds. New York: Springer Verlag. DOI= http://www.scs.gmu.edu/~dduong/Estimating_Impact_Corruption.pdf.

[5] Duong, Deborah. 2012. "Modeling the Effects of International Interventions with Nexus Network Learner." *AAAI Spring Symposium*. DOI= http://www.scs.gmu.edu/~dduong/Aaai12Nexus.pdf.

[6] Duong, Deborah, Jerry Pearman, and Christopher Bladon. 2013."The Nexus Cognitive Agent-Based Model: Coevolution For Valid Computational Social Modeling" *GECCO conference*. DOI = http://www.scs.gmu.edu/~dduong/GECCONexus.pdf .

ODD TEMPLATE FOR THE NEXUS COGNITIVE AGENT MODEL

1. Purpose

Question: What is the purpose of the model?

Answer:

The purpose of Nexus is to expand from basic principles to richer representations of real world scenarios, specifically to model emergent social behavior in a way that matches real world data, through a technique called data absorption. The Nexus Cognitive Agent Model is a general modeling framework from which agents can induce strategies that come to correspond with other agent strategies to serve agent goals. As agents choose one another based on dynamic traits and behaviors, dynamic social networks form. Nexus has been used in military analysis as a model of popular support and as a model of corruption.

2. Entities, state variables, and scales

Questions: What kinds of entities are in the model? By what state variables, or attributes, are these entities characterized? What are the temporal and spatial resolutions and extents of the model?

Answer:

Agents in Nexus are role-based. They seek to gain partners in corresponding roles that they qualify for. Different roles can be input in different networks according to function, for example, an agent can be part of both a kin network and a trade network. The agents in Nexus are general: their attributes, static or learned, are described in the input to the program, which consists of Bayesian Networks to describe the frequency of attributes in relation to each other at initialization, and ontologies to describe rules by which the agents may participate in network partner

choosing and transactions. The input also described which of these attributes are subject to learning with agent's individual Bayesian Optimization Algorithms (BOA), based on transactions that increase utility according to cultural goals (also described in the input ontology). Space can be represented by a separation of entitys from transaction according to a location attribute. Behaviors can be timed with a simulation clock, if a sense of time is desired in a model.

3. Process overview and scheduling

Questions: Who (i.e., what entity) does what, and in what order? When are state variables updated? How is time modeled, as discrete steps or as a continuum over which both continuous processes and discrete events can occur? Except for very simple schedules, one should use pseudo-code to describe the schedule in every detail, so that the model can be re-implemented from this code. Ideally, the pseudo-code corresponds fully to the actual code used in the program implementing the ABM.

Answer:

Individual agents first choose partners, then participate in transactions, and in some iterations change their strategies to a chromosome in their BOA, and periodically "think" or have their BOA's reproduced based on what got them the most utility. The number of chromosomes, a parameter of the simulation, determine how often these events take place. Agents behave simultaneously, in random order. There is a higher level controller that takes care of periodic

4. Design concepts

Questions: There are eleven design concepts. Most of these were discussed extensively by Railsback (2001) and Grimm and Railsback (2005; Chapter. 5), and are summarized here via the following questions:

Basic principles. Which general concepts, theories, hypotheses, or modeling approaches are underlying the model's design? Explain

the relationship between these basic principles, the complexity expanded in this model, and the purpose of the study. How were they taken into account? Are they used at the level of submodels (e.g., decisions on land use, or foraging theory), or is their scope the system level (e.g., intermediate disturbance hypotheses)? Will the model provide insights about the basic principles themselves, i.e. their scope, their usefulness in real-world scenarios, validation, or modification (Grimm, 1999)? Does the model use new, or previously developed, theory for agent traits from which system dynamics emerge (e.g., 'individual-based theory' as described by Grimm and Railsback [2005; Grimm et al., 2005])?
Answer:

As every agent in Nexus has a BOA which is not seeded from other BOAs, and learns only according to the perception of the agent, and since the agents learn the world together and change each others utility, the general approach of Nexus is coevolution thorough autonomy. Coevolutionary processes not only explain the social institutions emergent in Nexus, but facilitate the absorption of real world data.

Emergence. What key results or outputs of the model are modeled as emerging from the adaptive traits, or *behaviors*, of individuals? In other words, *what* model results are expected to vary in complex and perhaps unpredictable ways when particular characteristics of individuals or their environment change? Are there other results that are more tightly imposed by model rules and hence less dependent on what individuals do, and hence 'built in' rather than emergent results?
Answer:

This varies according to scenario: for the corruption scenario, the emergent behaviors are the patterns of a corrupt society.

Adaptation. What adaptive traits do the individuals have? What rules do they have for making decisions or changing behavior in response to changes in themselves or their environment? Do these traits explicitly seek to increase some measure of individual

success regarding its objectives (e.g., “move to the cell providing fastest growth rate”, where growth is assumed to be an indicator of success; see the next concept)? Or do they instead simply cause individuals to reproduce observed behaviors (e.g., “go uphill 70% of the time”) that are implicitly assumed to indirectly convey success or fitness?
Answer:

This varies according to scenario, for the corruption scenario, the agents learned strategies of behavior such as bribing and stealing, as well as whether to tolerate someone who will steal from you, or network with someone who offers bribes.

Objectives. If adaptive traits explicitly act to increase some measure of the individual's success at meeting some objective, what exactly is that objective and how is it measured? When individuals make decisions by ranking alternatives, what criteria do they use? Some synonyms for ‘objectives’ are ‘fitness’ for organisms assumed to have adaptive traits evolved to provide reproductive success, ‘utility’ for economic reward in social models or simply ‘success criteria’. (Note that the objective of such agents as members of a team, social insects, organs—e.g., leaves—of an organism, or cells in a tissue, may not refer to themselves but to the team, colony or organism of which they are a part.)
Answer:

This varies according to scenario, in the corruption scenario the goals are cultural based, and take the form of transactions that are desirable, such as, the number of times the maternal grandmother eats, for a matrilocal tribe.

Learning. Many individuals or agents (but also organizations and institutions) change their adaptive traits over time as a consequence of their experience? If so, how?
Answer:

Each agent has a BOA which tells it what behaviors to try out, and whether these behaviors succeed depends on other agents adapting corresponding behaviors and being chosen for network relations. For example, an agent may learn to bribe but its success depends on whether its network partners have learned to accept bribes.

Prediction. Prediction is fundamental to successful decision-making; if an agent's adaptive traits or learning procedures are based on estimating future consequences of decisions, how do agents predict the future conditions (either environmental or internal) they will experience? If appropriate, what internal models are agents assumed to use to estimate future conditions or consequences of their decisions? What tacit or hidden predictions are implied in these internal model assumptions?
Answer:

Each agent makes tacit predictions of its utility based on past experience through its own BOA.

Sensing. What internal and environmental state variables are individuals assumed to sense and consider in their decisions? What state variables of which other individuals and entities can an individual perceive; for example, signals that another individual may intentionally or unintentionally send? Sensing is often assumed to be local, but can happen through networks or can even be assumed to be global (e.g., a forager on one site sensing the resource levels of all other sites it could move to). If agents sense each other through social networks, is the structure of the network imposed or emergent? Are the mechanisms by which agents obtain information modeled explicitly, or are individuals simply assumed to know these variables?
Answer:

Agents can sense each others outward demographic characteristics and keep track of behaviors they have seen or that their network partners tell them other agents in their network have. Whether a network partner tells is probabilistic and based on role.

Interaction. What kinds of interactions among agents are assumed? Are there direct interactions in *which* individuals encounter and affect others, or are interactions indirect, e.g., via competition for a mediating resource? If the interactions involve communication, how are such communications represented?
Answer:

Agents have general transactions with each other, in which money may or may not be traded. Accounts are kept track of as well as appropriate role partners to engage in the exchange of money for services with.

Stochasticity. What processes are modeled by assuming they are random or partly random? Is stochasticity used, for example, to reproduce variability in processes for which it is unimportant to *model* the actual causes of the variability? Is it used to cause model events or behaviors to occur with a specified frequency?
Answer:

Bayesian networks control the frequency traits at initialization and of behaviors.

Collectives. Do the individuals form or belong to aggregations that affect, and are affected by, the individuals? Such collectives can be an important intermediate level of organization in an ABM; examples include social groups, fish schools and bird flocks, and human networks and organizations. How are collectives represented? Is a particular collective an emergent property of the *individuals*, such as a flock of birds that assembles as a result of individual behaviors, or is the collective simply a definition by the modeler, such as the set of individuals with certain properties, defined as a separate *kind* of entity with its own state variables and traits?
Answer:

Social Networks form dynamically and may be analyzed to describe entities.

Observation. What data are collected from the ABM for testing, understanding, and analyzing it, and how and when are they collected? Are all output data freely used, or are only certain data *sampled* and used, to imitate what can be observed in an empirical study ("Virtual Ecologist" approach; Zurell et al., 2010)?
Answer:

Which data and how often it is output is defined by the input parameters. It is output in the form of a bayesian network, a log of transactions, and the contents in the BOA of the individual agents.

5. Initialization

Questions: What is the initial state of the model world, i.e., at time *t* = 0 of a simulation run? In detail, how many entities of what type are there initially, and what are the exact values of their state variables (or how were they set stochastically)? Is initialization always the same, or is it allowed to vary among simulations? Are the initial values chosen arbitrarily or based on data? References to those data should be provided.

Answer:

Initial values are based on real world data, and vary from run to run, input in the form of a Bayesian Network.

6. Input data

Question: Does the model use input from external sources such as data files or other models to represent processes that change over time?

Answer:

Nexus uses a Bayesian network to represent agent attributes as they exist statistically in the real world, and ontologies that have cultural rules. There is also an option to input specific individuals. In the corruption scenario, data from the Congo was used.

7. Submodels

Questions: What, in detail, are the submodels that represent the processes listed in 'Process overview and scheduling'? What are the model parameters, their dimensions, and reference values? How were submodels designed or chosen, and how were they parameterized and then tested?

Answer:

Nexus process overview and scheduling parameters are inputs to the program. In the onotology that describes the cultural behaviors of agents, one may note behavior frequency, so that it is scheduled on the simulation clock.