Agent-based Modeling as an Evaluation Methodology

Jonathan A. Morell [1]
H. Van Dyke Parunak [2]
Kirk Knestis [3]
Melanie Hwalek [4]

Funding: The Faster Forward Fund provided partial funds for this research.

1- 4.669 Evaluation and Planning
Jamorell@jamorell.com
2 ABC Research
van.parunak@gmail.com
3) Evaluand LLC (Reston, VA)
kirk@evaluand.com
4) SPEC Associates
mhwalek@specassociates.org

Abstract

We present two scenarios of evaluation logic models cast as executable agent-based models (ABM). One scenario is the adoption of a therapeutic innovation. The second is a program designed to facilitate student's transition from community colleges to four-year universities. ABMs rely on the rules that govern the behavior of individual members of a group. This modeling method stands in contrast to equation-based models that are made up of system-level constructs. As an example of how ABMs and equation-based models differ, consider a model of the adoption of an innovative math curriculum by schools. An equation-based model might comprise schools, teacher training programs, set amounts of funding for teacher training, and equations governing relationships among these elements. An ABM would assign decision rules to schools who would then observe aspects of their environment – perceived advantage of the new program, funding, and number of other schools using the new program; and based on those observations, make a decision as to whether or not to implement the new math curriculum. Neither modeling method is better than the other, but each provides knowledge that the other cannot. We have four objectives. One is to show how logic models can be cast as ABMs. The second is to convey an understanding of how agent-based complex system behavior helps explain the operation of programs and their outcomes. The third is to interest a small number of evaluators into acquiring expertise in the construction of ABMs, and by so doing, providing a core of expertise to the evaluation community. Finally, we hope to convince our readers that evaluation will be better off if evaluators appreciated why thinking in agent-based terms can enrich the work we do.

Agent-based Modeling as an Evaluation Methodology

In this paper we present the results of two agent-based modeling (ABM) exercises (Morell & Parunak, 2015; Morell et al., 2016a, 2016b, 2016c; Parunak & Morell, 2014). In both, a traditional logic model was cast in ABM form in order to ascertain whether the model output would provide novel worthwhile understanding of program behavior and outcome. By reporting these results, we hope to inspire a few evaluators to form a core of expertise for the use of ABM in the evaluation community . Also, while we do not all need to be experts, evaluation will be better off if evaluators understood when and how ABM can enhance understanding about the interventions they evaluate.

Agent-based and Equation-based Modeling

Equation-based models are governed by mathematical expressions of how system-level components behave with respect to one another. As an example, such a model might show how a new diabetes control program is adopted across neighborhood health clinics as those clinics are supplied with training from state health departments. The health clinics, the diabetes control program, the training, and the state health departments are all system-level elements, and equations would govern interactions among them, e.g., how much training is supplied to how many clinics over a set period of time. As another example, consider a model of the adoption of an innovative math curriculum by schools. An equation-based model might comprise schools, teacher training programs, set amounts of funding for teacher training, all governed by equations that express relationships among these elements.

Sometimes such models can take a form that is familiar to evaluators – the stock and flow relationships that are so important in systems thinking (Sterman, 2001). Sometimes such modeling can involve differential equations applied to any number of constructs and scenarios (Daun et al., 2008). But there are always system-level components and equations to describe relationships among them.

The logic in an ABM is different. To take the math curriculum example, an ABM would assign decision rules to schools who would then observe aspects of their environment, e.g., perceived advantage of the new program, funding, and number of other schools using the new program; and based on those observations, make a decision as to whether or not to implement the new math curriculum.

What do Agent-based Models Add to Our Understanding?

Agent-based models simulate behavior of individual agents in order to study emergent phenomena at the level of the community (Šešelja, 2023). An "agent" can be any entity that can "sense its environment" and make "rules-based decisions" as to how to act based on what the agent senses. "Sense its environment" and "rules-based decisions" are put in quotes because in the domain of evaluation it is all too easy to think of agents as people. In fact, an agent can be any entity that can act *as if* it senses its environment and makes rules-based decisions. Thus an agent can be a clique of kids in a school, the school itself, a philanthropic organization, a department within a city government, a county government, and so on. In ABM, an important part of an individual agent's environment is the other agents that surround it. To continue the previous example, an equation-based model might contain an equation describing the rate at which the state health department provides training. An agent-based model would identify whether the state health department "decided" to provide training.

ABM provides three windows on a system's behavior that would not exist with other evaluation methods.

- In complex systems, local change can alter the trajectory of a system. ABM can reveal whether such local change occurs, where it occurs, whether the change does indeed alter the system (most of the time it will not), and if it does, what are the consequences of that alteration.
- Developing an ABM forces the modelers to attend to research, theory, and best judgement that identifies who and what the agents should be. The truer these choices are to reality, the better the model will be.
- ABMs are able to plot how individual agents and subgroups of agents act over time. This is because modelers have to specify the rules that agents will follow. The truer those rules are to reality, the better the model will be.

We will leave it to others to debate the epistemology of these models – whether they depict how the world really works, or whether they just provide useful descriptions to help us reason about phenomena of interest. The latter was enough to motivate us to undertake this research.

Research Plan

The research we report here proceeded through two stages. The first was to construct a simple artificial model that depicted a believable evaluation scenario, and to determine whether modeling it as an ABM provided novel insight that was applicable to evaluation. If Phase One provided enough intriguing results, we would proceed to the second phase – finding a real-world test case of a program evaluation.

Preliminary Test Model – Acceptance of a Treatment Innovation

We formulated a simple logic model unrelated to any actual program, but which seemed plausible with respect to a common topic in evaluation – the acceptance of an innovation in practice by therapists provided with training in its use (Figure 1) (Morell & Parunak, 2015; Parunak & Morell, 2014).

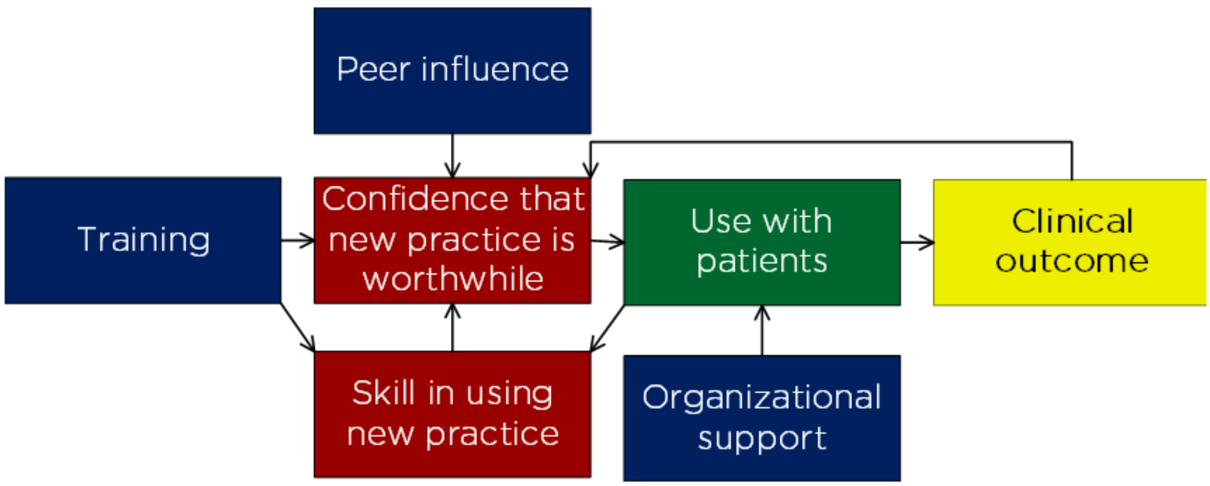

Figure 1: Practice Model - Therapist Acceptance of an Innovation

Figure 1 has four types of elements. The first are psychological characteristics of the therapists themselves (red in the figure). If they are skilled and confident in the new practice, they will use it with patients (green in the model). If they use the innovation, the therapists will observe a clinical outcome (yellow). Use increases skill and outcome affects confidence. Finally, the model considers outside influences affecting the therapists – peer influence, training, and organizational support (blue in the model). Peer influence and training have a direct impact on confidence and skill. Organizational support affects use with patients.

Then we used NetLogo[1] to cast the model as an ABM simulation. Controls for the model

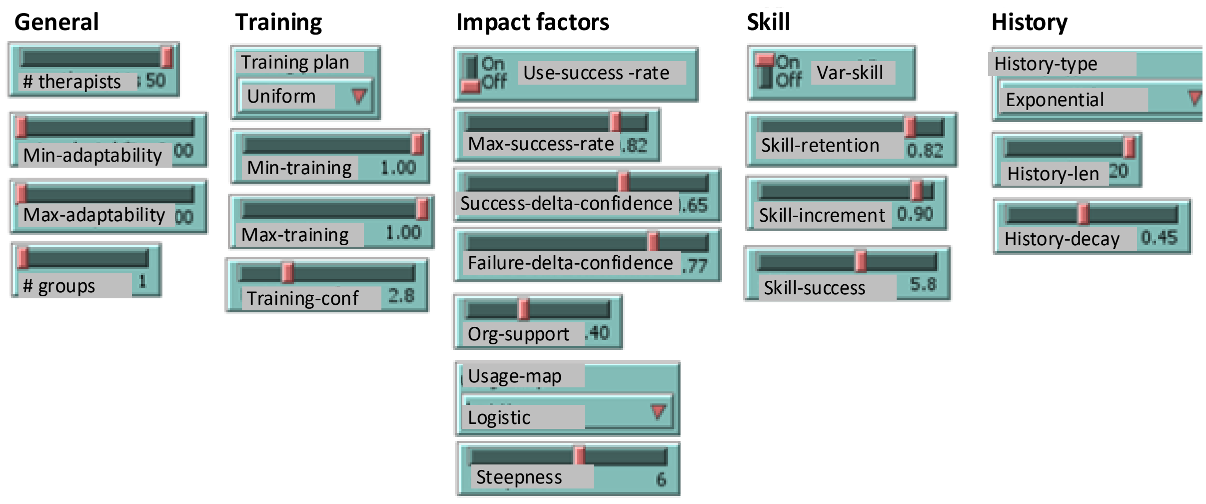

Figure 2 Net Logo Interface Innovation Adoption Model

are shown in Figure 2.

Two outputs of the model are graphs that would be familiar to any evaluator – therapists' confidence in the new therapy over successive iterations of the model, and percent adoption of the new therapy, also over successive iterations of the model. A third output is unique to an ABM simulation. It shows each therapist's adoption and confidence for each iteration of the model.

How does the model work? Fifty therapists are dropped into cyberworld, each starting with different (randomly assigned) levels of skill, organizational support, and so on. A value is set for the extent to which the new therapy is an improvement over traditional practice. A patient walks in the door. The therapist makes a decision. Use the new therapy or not? Observe the

[1] https://ccl.northwestern.edu/netlogo/

results. That observation, made by all fifty therapists, will affect levels of confidence, skill, and social support for each therapist and averages for the group as a whole. When the second patient walks in the door, the new environment of levels of support, confidence, and skill will determine each therapist's choice about whether or not to use the new therapy. And so it goes, round after round. (Watching the model run is the best way to appreciate insight that can derive from the ABM.[2])

Figure 3 shows the arrangement of therapists for two slices of time.[3] (Not every therapist is visible because their positions overlap.) Note the therapist icon surrounded by a circle. The circle identifies a single specified therapist that is tracked through each iteration of the model.

While Figure 3 shows only two slices in time, the pattern is consistent throughout. Over almost all the model's iterations, therapists' adoption of the new therapy is distributed relatively linearly (the variation along the Y axis is fairly uniform). In contrast, there is a "U" shaped unction with respect to confidence in the new technique. There are clearly high and low confidence groups of therapists.

---

2 https://www.youtube.com/watch?v=TW1C-K4R9ao&t=6s

3 In the graph, "confidence" ranges from zero to one. % implementation is the proportion of trials on which the therapist has adopted the innovation. In each iteration, we compute a threshold in (0, 1) based on individual confidence and organizational support, then generated a random number < 1, and if the random # is less than the threshold, we record an adoption. The y coordinate of a therapist is the % of trials in which the therapist adopted the innovation.

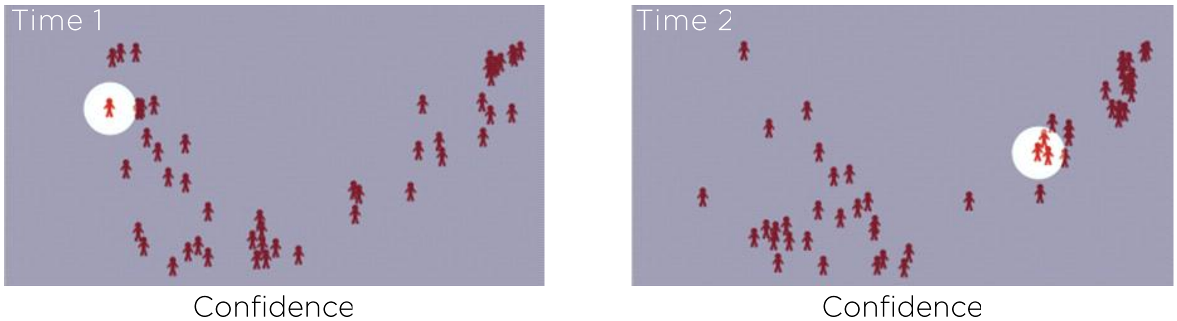

Figure 3: Illustration of Unique Ability of an ABM Model to Track Agents Over Time

What is particularly interesting, and which can only be revealed with an ABM, is how an individual therapist can jump back and forth between having high and low confidence at each iteration of the model. (The highlighted circle shows the same therapist at two successive iterations.) This finding can lead to interesting discussions. For instance, are we observing the behavior of therapists using an innovative therapy under this particular set of conditions as depicted in the model, or is this a general dynamic in mental health therapeutic environments?

Real-world Evaluation – Arizona General Education Curriculum

Our real-world test case was an evaluation of the Arizona General Education Curriculum (AGEC)[4]. AGEC is designed to facilitate college attendance by eliminating friction in the transition from community college to universities. As their website states:

> The Arizona public community colleges and universities have agreed upon a common structure for a general education core. This curriculum provides students attending any

[4] https://catalog.asu.edu/agec

Arizona public community college with the opportunity to build a general education program that is transferable to any other state institution without loss of credit.

A key element of the AGEC evaluation was based on the program theory that students who transitioned to four-year colleges through the Arizona community colleges' general education core would attain bachelor's degrees at a higher rate than students without certification, theoretically because of two elements: 1) they were better prepared and 2) because more of their credits would transfer. The logic of the program is depicted in Figure 4. The thickness of the arrows expresses magnitudes of if →then relationships that are baked into the program theory. AGEC students do and do not graduate from university, but AGEC preparation makes their graduation more likely. Non-AGEC students do and do not graduate, but for them, graduation is less likely.

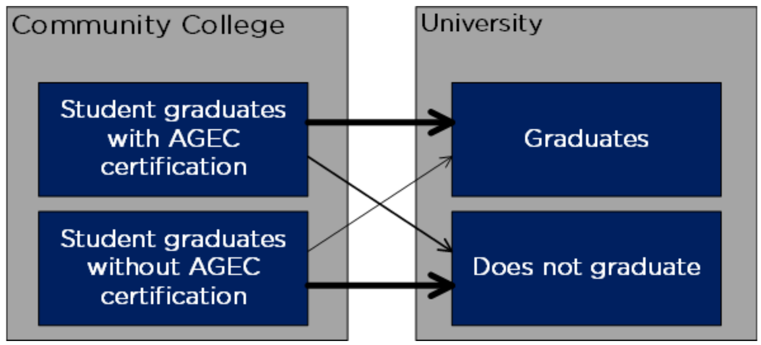

Figure 4: Core AGEC Logic

Our challenge was to produce a model that met five criteria. 1) It must be executable as an ABM. 2) It must output data on graduation rates. 3) It must be consistent with, at least in spirit, to the logic of the AGEC program. 4) It must reflect an "evaluation mindset", i.e. the model must be recognizable as an evaluation logic model, and conducive to the kind of evaluation designs that we commonly use. 5) The model must be practical as a test within the limited time and resources we had available for the research. The model we developed is shown in Figure 5.

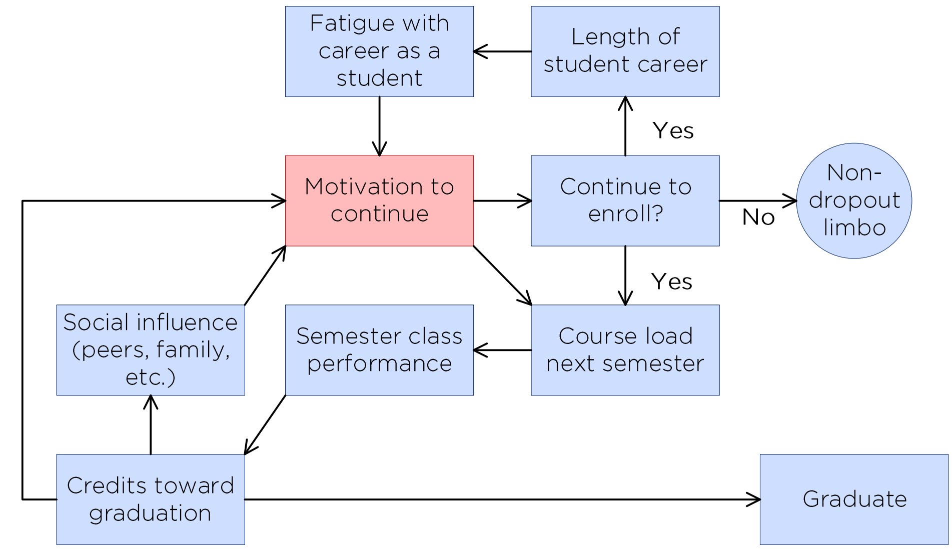

Figure 5: AGEC ABM Model

Relationships in the model are:

- The model is centered on a student's motivation to continue. Motivation leads to a decision to enroll for the next semester, and to a specific course load.
- Enrollment increases the length of the student's academic career. Length leads to fatigue with life as a student.
- Fatigue with one's student career negatively affects motivation to continue.
- Courses lead to academic performance
- Performance leads to credits toward graduation.
- If the student amasses sufficient credits, they graduate. Otherwise, performance leads to social influence from peers and family.
- Social influence affects motivation.
- There is a direct relationship between credits earned and motivation to continue.

When transformed into an ABM, Figure 5 becomes Figure 7, Figure 6 and Figure 8. Figure 7 shows the controls for the model.[5]

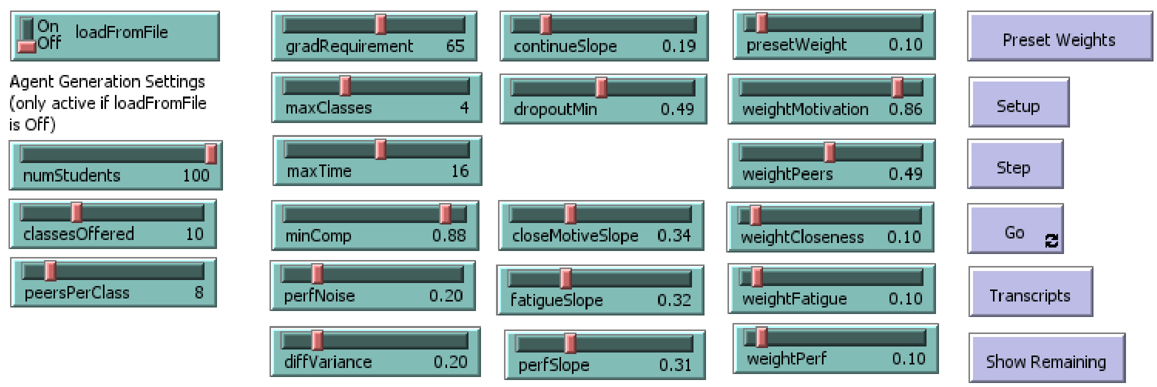

Figure 7 AGEC Model ABM Interface

Figure 6 shows traditional variables of a type we are all familiar with, e.g. percent graduated, mean number of courses, and motivation. What makes this view unfamiliar is that it shows those values as they change continually over time (semesters), with each iteration of the model.

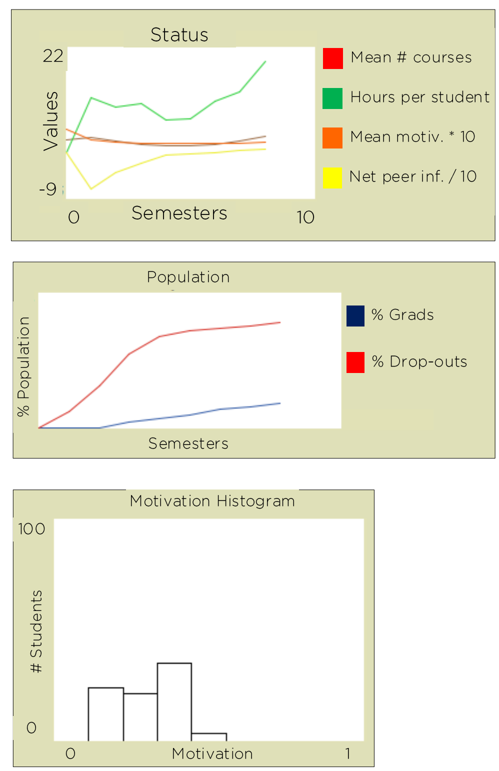

Figure 6 Parameter Readout AGEC

Figure 8 shows us six configurations of each students' movements as the model runs. It gives us a movie of how the population of different types of

---

[5] Code for the model is available upon request to the authors.

students evolve as they make decisions about their course activities. In the model, green icons are students who have AGEC certification, red for students who do not have AGEC certification.[6] (For a better sense of how these configurations develop over time, it is necessary to view the video.[7])

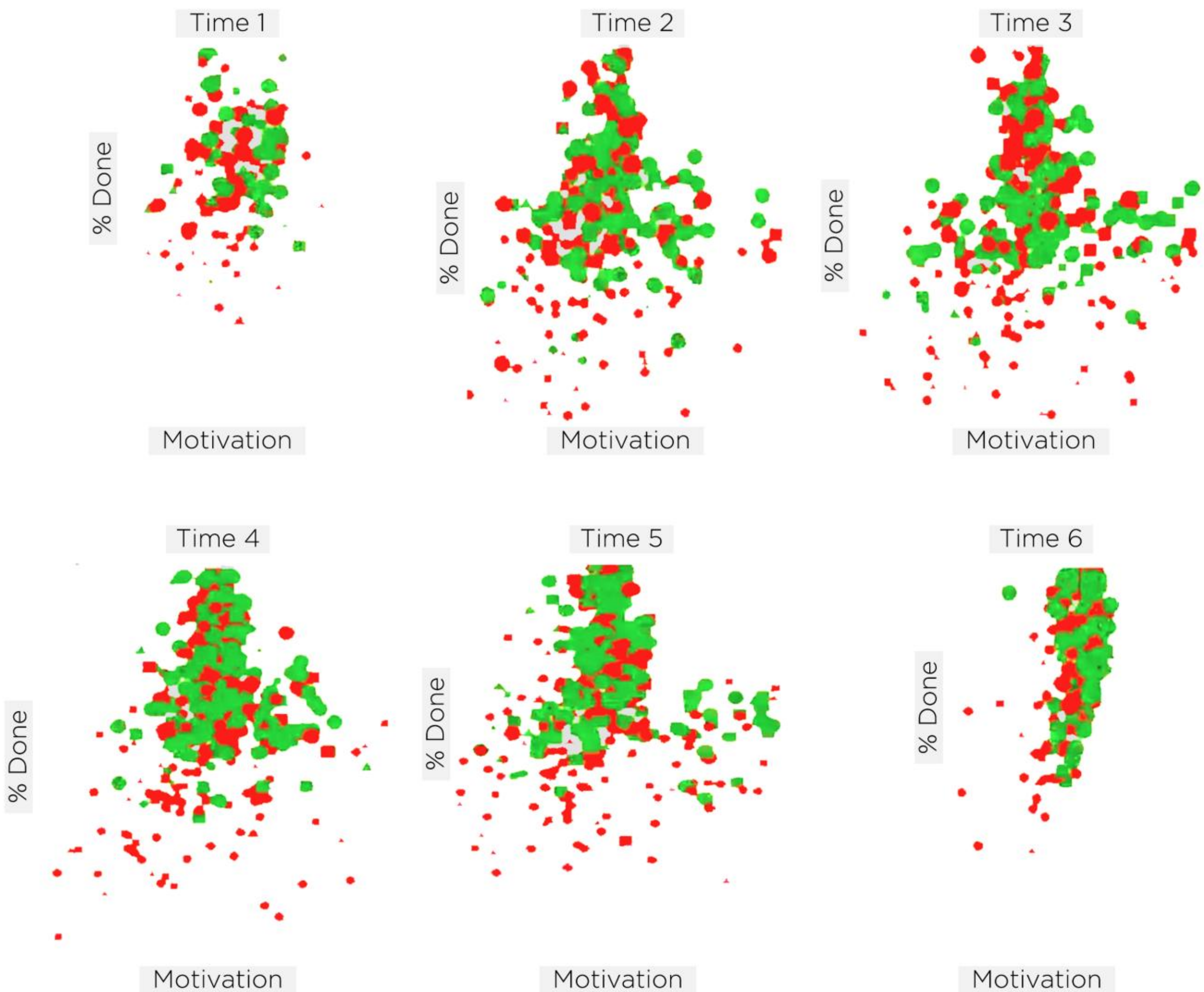

Figure 8 Agent Status at Six Times in a Model Run

[7] Part 1 https://www.youtube.com/watch?v=jWDKSHQs0rA 6:51
Part 2 https://www.youtube.com/watch?v=aoO1lIEcMzk 20:18
Part 3 https://www.youtube.com/watch?v=8bSOO3lOvwI 19:57

Several observations are worth noting. First, soon after the model begins, most students are tightly lumped together. (There are more students than visible icons because so many students' positions overlap.) Second, as the model runs, students' positions spread out widely with respect to motivation, while success in pursuing a degree progresses slowly. Third, students with AGEC certification (green icons) gravitate toward the top end of the percent done scale, while students who lack AGEC certification (red icons) gravitate toward the lower ends of the percent done scale. Finally, the spread in motivation narrows toward the end of the model's run even though there is a discernable difference in progress toward graduation. The display shows the evolution of student behavior that can only be provided with an ABM. This is because ABMs do not focus on the means and standard deviations of group behavior. They focus on individual agents and patterns of their behavior over time.

A traditional group-parameter analysis provides a complementary understanding of student behavior (Figure 9). Here, we plot students' average motivation over twenty semesters.[8]

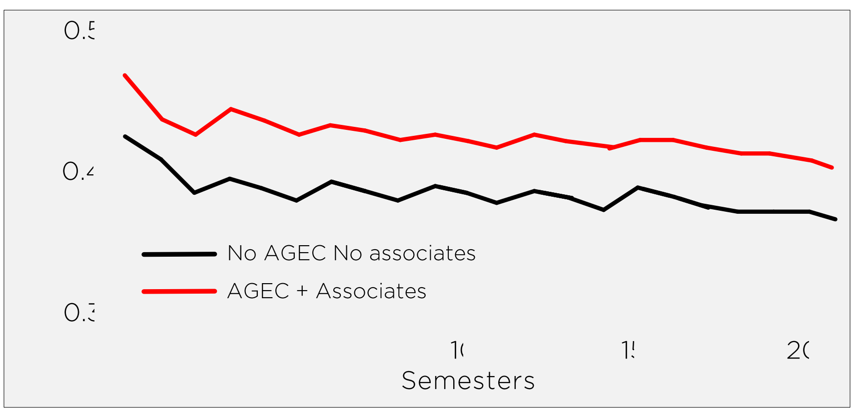

Figure 9 Motivation Over Semesters

[8] Data for four groups (1) no AGEC no Associates, 2) no AGEC yes Associates, 3) Yes AGEC no Associates, and Yes AGEC yes Associates, show a similar, but muddled relationship. Combining the data into AGEC versus no AGEC smooths out the noise and makes the pattern more obvious.

It is clear that students with a combination of AGEC certification and an Associate's degree, maintain higher motivation than students with neither AGEC certification nor an Associate's degree.

Figure 8 and Figure 9 and are not perfectly compatible because Figure 8 plots motivation (X axis) against closeness to graduation ("done" – Y axis), while the other plots length of student career (semesters – X axis) with respect to motivation (Y axis). Still, the stories provide complementary understandings into the role of motivation.

Figure 8 expresses the pattern in terms of the individual behavior of each student. We can follow people. We cannot, however, perceive the overall behavior of the *group* of students.

Using the Models

We do not claim that ABM is better than equation-based models, but we do claim that it is different and valuable. What makes it different is that it provides a view of system behavior that would be invisible by other means. We do not use the term "invisible by other means" to connote a more accurate or powerful way of detecting what was always known. By analogy, placing an optical telescope on a mountaintop to reduce atmospheric distortion provides more information on classes of astronomical phenomena that were always known to exist. We can see those things better, but not differently. In contrast, imagine the invention of the radio telescope. Those instruments provided knowledge about astronomical phenomena that were previously unknown. So it is with ABM. Moving from understanding based on means and variances and adopting an agent-based orientation shows us a class of behavior that was previously unknown. Moreover, as we hope we have demonstrated, it is a class of behavior that can inform program theory and shape decisions about evaluation design and methodology.

To be sure, both models presented above generate patterns that raise questions about the fit between the model and reality. We cannot escape asking whether the output of the models makes sense. This is as it should be because of the numerous assumptions that are made. There are assumptions in the static models from which the ABMs are built. And once the models are built, there are assumptions about each control that can be set (Figure 2 and Figure 7).

If the patterns of agent behavior generated by an ABM are so anomalous, why go to the trouble of building the models? In many fields the answer is that eventually the models can be brought to the point of accurate prediction, at least within a specified time horizon (Orrell, 2007). But given the resources needed to build such models, and the state of our knowledge about program effects, prediction does not seem like a prime reason to do modeling in evaluation. Rather, we should build models to generate outputs to question so that we can modify the model, run it again, and question the output again – over and over, tinkering with the model controls, and changing the architecture of the static model on which the executable model is built (Figure 5). Our goal should be to reach a point where the model's output is plausible, or better still, comports with what is known about similar programs. Even if our ABMs never get to the prediction stage, the act of questioning and revising the model can help us explain why a program may operate as it does, and that understanding is valuable indeed. Ideally, model building and revision will repeat at regular intervals over the course of an evaluation.

References

Bendassolli, P. F. (2013, 01/29). Theory Building in Qualitative Research: Reconsidering the Problem of Induction. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 14(1). https://doi.org/10.17169/fqs-14.1.1851

Bryman, A., Stephens, M., & Campo, C. à. (1996, 1996/09/01/). The importance of context: Qualitative research and the study of leadership. The Leadership Quarterly, 7(3), 353-370. https://doi.org/https://doi.org/10.1016/S1048-9843(96)90025-9

Cragun, D., Pal, T., Vadaparampil, S. T., Baldwin, J., Hampel, H., & DeBate, R. D. (2016, Jul). Qualitative Comparative Analysis: A Hybrid Method for Identifying Factors Associated with Program Effectiveness. J Mix Methods Res, 10(3), 251-272. https://doi.org/10.1177/1558689815572023

Daun, S., Rubin, J., Vodovotz, Y., & Clermont, G. (2008, 2008/12/01/). Equation-based models of dynamic biological systems. Journal of Critical Care, 23(4), 585-594. https://doi.org/https://doi.org/10.1016/j.jcrc.2008.02.003

Franek, J., & Kashi, K. (2014). A Review and Critique of MADM Methods And Applications in Business and Management. International Journal of the Analytic Hierarchy Process, 6(2). https://doi.org/https://doi.org/10.13033/ijahp.v6i2.254

Green, A. E., Fettes, D. L., & Aarons, G. A. (2012, Oct). A concept mapping approach to guide and understand dissemination and implementation. J Behav Health Serv Res, 39(4), 362-373. https://doi.org/10.1007/s11414-012-9291-1

Hacking, I. (1990). The Taming of Chance. Cambridge University PressChapter 10 quote taken from: A.D. Poisson, P.L Dulong, D.J. Larrey and F.L. Double Weekly reports of the Academy of Sciences Sessions 1 1835, pp 167 - 177.

Morell, J., Parunak, H. V. D., Hwalek, M., & Knestis, K. (2016). Combining Agent Based Modeling and Traditional Evaluation. https://www.youtube.com/watch?v=jWDKSHQs0rA

Morell, J. A. (2024). Characteristics of Sustainability - A Complexity-based View to Inform Evaluation (In press. Draft available upon request.). Evaluation and Program Planning.

Morell, J. A. (2024). A Complexity-based Plan for Evaluating Transformation. Journal of Multidisciplinary Evaluation. https://journals.sfu.ca/jmde/index.php/jmde_1/article/view/867

Morell, J. A. (2024). Early Childhood Development Action Network Master Class in Systems Thinking: Complex System Behaviors that Can Help Us Plan and Evaluate https://evaluationuncertainty.com/2024/04/23/early-childhood-development-action-network-master-class-in-systems-thinking-complex-system-behaviors-that-can-help-us-plan-and-evaluate/

Morell, J. A., Parunak, H. V. D., Hwalek, M., & Knestis, K. (2016a). Part 1 Combining Agent Based Modeling and Traditional Evaluation. https://www.youtube.com/watch?v=jWDKSHQs0rA

Morell, J. A., Parunak, H. V. D., Hwalek, M., & Knestis, K. (2016b). Part 2 Combining Agent Based Modeling and Traditional Evaluation [Movie]. https://www.youtube.com/watch?v=aoO1lIEcMzk

Šešelja, D. (2023). Agent-Based Modeling in the Philosophy of Science (Edward N. Zalta & N. Uri, Eds. The Stanford Encyclopedia of Philosophy https://plato.stanford.edu/archives/win2023/entries/agent-modeling-philscience/

Author Biographies

Dr. Jonathan (Jonny) Morell is founder of 4.669… Evaluation and Planning. He does evaluation, research and writing on the application of complexity science to evaluation. Jonny believes that models, methodology and data interpretation should recognize the behavior of complex systems, but that evaluations should be as simple and straightforward as possible. His blog is Evaluation Uncertainty: Surprises in Programs and Their Evaluations,

Dr. H. Van Dyke "Van" Parunak is a senior scientist at Parallax Advanced Research, and president of ABC Research LLC. His research experience includes linguistics, biology, complex adaptive systems, and software development, and recently has focused on causal formalisms and simulation platforms for socio-technical systems.

Dr. Kirk Knestis is founding Principal of Evaluand LLC, a research consultancy local to Washington DC. He has been a professional evaluator and researcher for more than 20 years, having previously been a business owner, K-12 STEM and arts educator, and university faculty member. A content expert in STEM and workforce education, he specializes in mixed-method evaluations; research and development to test and improve education and social services innovations; and the design of theory-based studies to understand implementation and outcomes of complex, multi-level, and multi-site change innovations.

Dr. Melanie Hwalek is CEO of SPEC Associates and Assistant Professor of Psychology at Michigan State University where she teaches the graduate level course in Evaluation Management. She has been directing program evaluations on a myriad of topics for more than 40 years.