Recursive Agent-Based Models in which Certain Agents Utilize Their Own Agent Models

Robert Axtell[1,2] and Scott D. Eldridge[1]

[1] George Mason University, Fairfax VA 22030, USA
[2] Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe, NM 87501 USA
{rax222,seldrid}gmu.edu

Abstract. One motivation for using agent-based systems as models of social phenomena is that they may provide higher verisimilitude than other approaches, e.g., mathematical models written in terms of representative agents or more black box machine learning techniques. In such agent systems, individuals make decisions using relatively simple mathematical models, based on the state of their environment. In certain circumstances it may be beneficial for at least some agents to use their own agent models in order to figure out how to behave. Here we demonstrate this idea, in principle and through three examples. We begin by discussing models in which each person being represented has its own agent-based model, one level down. We go on to discuss such models nested L levels deep. An example implementation of the Schelling segregation model illustrates the main ideas and demonstrates that agents who employ their own agent-based model can arrive at different decisions than they would otherwise. A second example involves the so-called El Farol model in which each agent has available to it a variety of decision rules, one of which is an agent model, and we test how often the agent model outperforms other models. We relate our work to multi-agent systems models in which agents build models of one or a few other agents, to cognitive science research concerning the theory of mind, to efforts in game theory on k-level cognition, and to the so-called 'Lucas critique' in economics.. Overall, we find significant promise in our approach and speculate about the kinds of agent systems that might benefit from the methodology we outline.

Keywords: recursive multi-agent systems, nested agent-based models, agents who use agent-based models

1 Introduction

Imagine driving up to a 4-way intersection, governed by stop signs. Your intention is to turn left and you arrive approximately contemporaneously with a vehicle facing you that has signalled its intention to turn its left (your right). Also pulling up to the intersection, on your right, is another vehicle with no turn signals flashing, ostensibly it will proceed straight, but coming to a full stop a moment after you and the car across from you. Which vehicle moves first, which next? While answers to these questions may depend on local driving norms, with the ‘correct’ order of access to the intersection being somewhat different in distinct jurisdictions, one answer that is typical of a variety of locales goes as follows:

1. The vehicle to the right that has stopped last has the least priority and must wait for the other two;

2. The two vehicles that are each turning left have equal priority and, if they are able to move simultaneously without interfering with one another, do so; if turning left at the same time would cause interference then the movement order is indeterminate;

3. After the two vehicles have left the intersection the third one goes straight.

Other possibilities certainly exist, such as the vehicle to the right having the right-of-way to move first if its delay in stopping relative to the other vehicles was very short, since it is going straight while the others are turning. How short a lag moves one solution to the other? How do individual drivers perceive the lag, and how do they assess how other drivers assess it?

These are not questions that are easily dealt with using conventional decision theory or game theory. While the situation described is a 3 person game, issues of perception and the perception of perceptions may be decisive. While attempts to address such strategic situations exist in game theory, the specific circumstance described involves matters (timing, perception of timing) that are somewhat ‘below’ (i.e., more ‘micro’ than, perhaps ‘nano’ in comparison to) the usual specifications of strategic environments.

It turns out that in other fields there are methodologies that might, in principle, address such situations. In cognitive science one of these is known as ‘simulation theory’ and considers agents who ‘run’ simulations in their head of the specific circumstances in which they find themselves. The exact kinds of simulations to be used are not specified but seem to depend on the social context and may or may not be methodologically individualist. Another approach comes from behavioral psychology/economics and amounts to having agents follow heuristics. In the situation described above there are a variety of these that might be invoked. Such heuristics may not be methodologically individualist in character.

Here we investigate models of social phenomena in which individuals use explicit models of other individuals. That is, we adopt a pure methodologically individualist (MI) position, as distinguished from conventional methodological individualism in which each individual agent has some mathematical representation of its social world that does not fully distinguish individual agents, i.e., is not MI. We apply this perspective to a variety of social and economic situations, finding that it can lead to behaviors at the individual level that aggregate into novel social outcomes. We distinguish shallow and deep methodological individualism and suggest that extant MI theories suffer from a kind of inconsistency that may make their conclusions not robust.

1.1 Agents and Methodological Individualism

Methodological individualists seek explanations of social phenomena grounded in the actions of the individuals who are active in the phenomena. Typically, they reject theories of social phenomena that abstract from the behavior of individual people, e.g., mathematical theories written in terms of group goals or behaviors. For example, a model of consumption that abstracts from the preferences of individual consumers is not MI.

The notion of methodological individualism has been a core tenet of certain branches of the social sciences since the late 19th C, implicitly in Menger [18], explicitly in Weber [21]. It arose partially as a reaction to certain fallacies of division in social contexts, such as when the goals or objectives of a particular group would seem to be furthered by certain kinds of behaviors on the parts of individuals who compose the group, yet when evaluating the incentives facing those individuals, different behaviors appear to be preferred by them. Today there are many formal examples of this, from so-called 'tragedy of the commons'-type situations to any game in which dominant strategies lead to Pareto-inferior outcomes. Perhaps more generally, there is a certain sense in which methodological individualism is a kind of reductionism applied to social science problems, in which overall social behavior is viewed as the resultant of individual behaviors, in much the same way that the temperature of a volume of gas can be derived from the motions of the individual gas particles. Indeed, some have argued that MI not merely a cornerstone concept of modern social science, but the core idea differentiating it from the natural sciences [8].

However, as an core tenet of some of the social sciences, methodological individualism has been regularly criticized, from a variety of social science disciplines and schools of thought (e.g., Little [15]). Specifically, sociologists have long argued that the decomposition of social groups into collections of individuals does violence to group-centric concepts such as group identity, group behavioral norms, and group incentives. Their perspective is that social constructs such as these so constrain individual actions as to be decisive in many contexts, making focus on individual preferences and actions of secondary or tertiary interest. However, certain sociologists (e.g., Coleman [7]) have argued that it is precisely these group conventions and regularities that should be explained by invoking the MI perspective. In particular, the emerging 'analytical sociology' tradition (e.g. [7]) seeks explanations for group-level phenomena in the interactions of individuals through explicit mechanisms. In demography, anthropology, politics, and even linguistics, similar debates have played out, regarding the tension between using top-down explanations of mass or group behavior based on social forces vs attempting to explain the social forces from the bottom-up via the actions of individuals. While today these arguments are in play in nearly each issue of contemporary social science journals, there is broad acceptance of the MI stance throughout wide swaths of the social sciences, increasingly buttressed by emerging ideas from complex systems (e.g., 'more is different' [1]) and the application of computing technologies (e.g., agent-based modeling (ABM), in which the actions of large numbers of individuals can be simulated.

In what follows we shall not directly engage these debates on the merits of methodological individualism, due to space constraints, but will attempt to extend the scope of the MI position, conceptually, by deepening its reach and, thereby, broadening its applicability, while demonstrating that the conventional form of methodological individualism is somewhat naïve.

1.2 Agents as Methodological Individualists

The desire to create working models of social phenomena that are MI often outstrips the ability to solve such models and so, in practice, a kind of middle ground is where many models live. For example, while there are hundreds of millions of consumers in the U.S., neither sufficient data exist on the actual tastes and endowments of these people to do anything like solve a huge number of equations to determine actual market-clearing prices and quantities for American markets. Rather, when policy-makers need to assess the effects of a change in some policy on U.S. consumption patterns, they resort to 'representative agent' models [13] for which the hope exists that, by understanding the incentives facing one or a few (types of) consumers, insight into the behavior of all consumers can be achieved. Clearly this hope may be fraught with error in a system as complex as a modern economy, but as the only way to maintain a MI stance, it has become a norm in economics research.

However, in MI models, what motivations are assigned to the individuals being modeled? That is, as a matter of principle, are the agents themselves in MI models MI? The short answer, with a couple of exceptions to be described in more detail below, is readily known to all who engage in mathematical or computational modeling of social phenomena: no, our agents are not themselves MI. Rather, our agents have some reduced form representation of their social world that is, at least largely, devoid of specifications of the individuals who make up their environment, and instead use some kind of simplified, mathematical representation of their circumstances. In essence, the agents in our models do what we deny is satisfactory behavior on our own part, i.e., representing the relevant social phenomena in a way that abstracts from the individuals present. Conventional MI models are inconsistent, in this sense—they do not employ (representations of) individuals who are themselves MI!

The way to repair this lacuna seems obvious enough: each individual in a model of social interactions needs to have, in its head, as it were, a model that is MI. This fix is easy to state in words but potentially challenging to realize in practice. What is hard, mathematically, at least in all but the most trivial circumstances, is to write down models of purposive (goal-directed) agents who use their own models of other agents in deciding how to behave. Precisely for the reason that representative agents are the tools of choice when the number of agents is large, solving models in which agents have to solve their own agent models appears to be a daunting task. However, computational agent systems typically do not involve 'solving' systems of equations, but rather simply let the agents interact through rules of behavior in order to reproduce social outcomes. Perhaps it is possible to make progress on models in which each agent has in its head its own ABM—i.e., nested agent models—by simply running such models with specific behavior rules in place.

2 Nested Agent Models *One Level Down*

Agent systems come in several different flavors, such as (1) when they are used to represent a specific social phenomenon, grounded in some amount of micro-data that are used to specify agent behaviors, (2) when they are used as purely theoretical devices, for studying how macro phenomena emerge from micro behaviors, and (3) as engines for dealing with some task that can be rendered computationally. Here we explore the application of agent technologies to situations in which each agent in an ABM of some social process uses an ABM of its own to make decisions. That is, we will have two 'layers' of ABMs, the topmost one, which we will usually refer to as level 0, representing the social system under study, and a large number of them one level down. While it might be the case that we have sufficient data to build models of this type that are grounded empirically, we will start by discussing purely theoretical models.

Consider the case of three vehicles mutually approaching an intersection governed by stop signs, as described in section 1 above. Now consider modeling this situation with an ABM in which each of the agents has at its disposal ABM of the other agents. That is, we create a three agent ABM involving all three vehicles, and then we give each of the drivers of these vehicles a two agent model of the other agents present; overall, one 3 agent ABM and three 2 agent ABMs, 4 ABMs in total involving 9 distinct software agents, all to model a situation involving 3 people. Now, in order for each agent to decide what to do, it runs its own ABM, based on its perceptions of the social circumstances and its models of the other agents. In essence, each of the three agents runs a 2 agent ABM locally and, based on the results produced, decides what to do. There are many possible outcomes, of course, depending on perceptions and so on. How this overall model might be realized in software will be the topic of the next section. Here we simply point out that, overall, of the 9 agents are needed for this model to run, several represent the same person, and it remains to be seen whether such a model can produce results different from what one would expect from traditional game theoretic or other formal treatments of the problem.

As another example of a two-layer ABM, consider Keynes' beauty contest, which is well-known in economics as a model of expectations formation and illustrates multi-level thinking. A page of N beautiful objects, say flowers, is shown to a group of people. Each person votes once for the most beautiful object, with the winning entry receiving V votes. Then, each of the V people who voted is given a $1/V$ chance of winning a prize. Keynes knew that the best strategy for any particular player of this game was *not* to vote for the object the player assessed to be most beautiful, but to vote for the one that the player believed most people would find most beautiful. Game theoretic models of so-called k-level rationality are a formal version this set-up, such that for any agent that looks ahead k moves, there always exists a $k+1$ strategy that outperforms it [6]. Computational models exist for computing a better strategy for any k for certain games [14].

One way to make progress with nested agent systems is reasonably clear from these two conceptual examples. Each agent in the topmost model simply runs its (imperfect) agent model in order to decide what to do—when to drive through the intersection or which object to vote for. That is, the way one 'solves' an "ABM of ABMs" is to simply run the lower level models as decision aids for the agents at the highest level. This is readily accomplished computationally, conditional on having plausible behavioral rules for the lower level agents, and involves no elaborate mathematics or solution procedures: just execute the models! Under what conditions will this methodology yield results that are similar to conventional models, whether MI or not? What are conditions in which multi-layer ABMs yield different results from corresponding one-level MI models? At this time we do not have a general theory of how or when this might happen so we illustrate some possibilities below.

2.1 Example 1: Residential Segregation

In the original checkerboard model of segregation due to Schelling [20], arguably the first true ABM in the social sciences, there are no equations solved, nor do agents attempt to model one another. Rather, agents who are unhappy in their current neighborhood, due to having too few neighbors who are like themselves, simply move to a location where there are enough neighbors to make them happy, given their fixed preferences. That is, 'solution' of the model involves running it, and under a wide variety of specifications of preferences and moving rules it results in the entire population living in neighborhoods in which each has a sufficient number of neighbors of the same type to satisfy its preferences. The contribution of the Schelling model to the study of segregation involves the fact that the realized number of like neighbors, on average, greatly exceeds the preferences of the agents, suggesting that individual preferences may play only a minor role in the high levels of residential segregation present in many parts of the world.

Of the dozens of implementations of this model in the public domain, including many that are part of libraries of some of the most widely-sued ABM software frameworks and packages (e.g., NetLogo [22], MASON [17], MESA, Hash.AI), it is uncommon to find one one in which the agents create models of their social environment. Typically, unhappy agents just move to a nearby site where they will be happy once they arrive there, making no inferences concerning how their new neighborhood will evolve, i.e., they are myopic. However, such agents have incentive to model in some detail the various destinations they consider moving to as it is not uncommon in Schelling-type models for agents to have to move multiple times due to situations in their new neighborhoods leading them to be unhappy. Indeed, before individual runs of such models settle down to all agents being happy, it can be the case that some agents move many times. It would be useful to unhappy agents if they had some plausible way to assess the long run desirability of a neighborhood and thus rank alternative destinations that look equivalent, myopically, but may not remain so over time. Typical implementations of the Schelling model do not do this, yet is seems clear that unhappy agents could benefit from each agent running an ABM of its alternative landing sites in order to assess the likely duration of its future occupancy at each.

We have created such a model for a landscape (50 x 50) in C++, populated by 2400 agents, each being one of two types and having preferences for having a certain fraction of its neighbors being of its same type. In this model there are typically about 100 agents who are unhappy after being given random initial locations, and they progressively move to locations that make them happy. Figure 1 plots, in blue, the histogram of the number of agents who make one or more moves due to being unhappy, averaged over 100 runs. It takes a few agents as many as 10 moves, acting myopically.

For this model we have created the ability for each unhappy to assess, before moving, the desirability of each prospective move location by running the model forward a few periods to learn how the neighborhood evolves. This involves checkpointing state information at the time the sites are assessed and restoring this before each unhappy agent's move. Overall, each unhappy agent makes its own idiosyncratic assessment of how local neighborhood will evolve. Running this nested agent model yields the red curve in figure 1. Starting out with approximately the same number of unhappy agents, the effect of running the nested models is to improve the ability of unhappy agents to find places where they will be happier longer. For the parameters used in the model that yields figure 1, essentially no unhappy agents move more than 6 times.

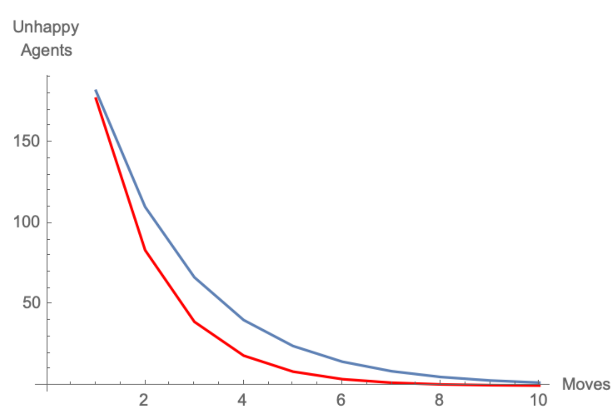

Fig. 1. Unhappy agents engage in fewer subsequent movements when they have their own agent model

We have experimented with many variations of this model and will report these in future work. Here, suffice it to say that the *possibility* of nested agent models to produce outcomes different from conventional implementations has been demonstrated.

2.2 Example 2: El Farol

A second nested ABM involves the El Farol Bar problem, a conceptual model of inductive reasoning under circumstances of limited information. Originally theorized by Brian Arthur [2], this problem supposes that 100 individuals consider attending live music on Thursday evenings at the El Farol bar in Santa Fe. With limited space available, each of the 100 individuals must independently decide whether they will attend on Thursday or stay home. The consensus view is that if fewer than 60 people attend on a given Thursday, the bar is not too crowded and the atmosphere is favorable for all to have fun. Each week the individuals predict whether fewer than 60 will attend, and they make a decision to go or stay home accordingly. Individuals don't know in advance who will attend, and the only predictive information available is the previous week's attendance count. Further, individuals are not allowed to collaborate nor communicate their intentions to others. After the Thursday night attendance is tallied each week, the individuals re-evaluate their prediction strategy and switch to a strategy that, in hindsight, was most accurate in prediction for that week. This process repeats for the next week, and so on. As Arthur points out, the collective solution for the group to achieve fewer than 60 people in attendance cannot be found through deductive means. Instead, the problem supposes that each individual will adopt a number of inductive strategies to best predict attendance each week.

A number of papers have considered this problem and have proposed modifications and extensions to improve the inductive predictive performance of the baseline El Farol model. For example, Wilensky and Rand [22] reinterpret the idea of inductive predictions in the form of a set of weights used to determine the relative importance of past predictions to suggest future predictions. Fogel et al. [1996] propose an evolutionary variation of El Farol, where agents adapt their prediction strategies over time to eliminate ones that are ineffectual. Baccan and Macedo [1999] suggest an approach using cognitive modeling to mimic the emotional state of participants as a scheme to formulate attendance predictions. There are numerous other adaptations and experimentations of this model.

Here we propose that the agents include a nested ABM as an additional inductive strategy to improve predictive performance. Theoretically, a Level 0 agent could recursively infer the attendance predictions of each of his or her 99 cohorts. By using those recursively inferred predictions as a decision factor in making his own attendance predictor, a Level 0 agent ought to improve its overall prediction accuracy. The expected net effect is that there should be a reduction in the overall variance of attendance because agents are able to realize a benefit of recursively inferring the predictions of their peers.

A baseline El Farol model was built in Java using the model specifications and predictive methods outlined by Arthur [2]. Agents are randomly endowed with three of six available heuristic functions, called predictors. The predictors forecast weekly attendance for each round using the following methods:

1) Same as the previous week (i.e., 1-week look back);

2) Attendance mirrored around the El Farol's capacity midpoint (i.e., 50 people) from last week's attendance;

3) Moving average of the last four weeks;
4) Linear regression of the past eight weeks, between 0 and 100;
5) Same as two-weeks ago (i.e., 2-week look back);
6) Same as five-weeks ago (i.e., 5-week look back).

The model execution starts when one-hundred Level 0 agents are instantiated and randomly assigned one of three available predictors to be their initial active predictor. At the beginning of each round (week), the activation sequence of the Level 0 agents is randomized. Each agent produces an attendance prediction based on its active predictor and makes their decision to attend. Once all agents are activated, the actual weekly attendance is tallied. Each agent then evaluates the prediction accuracy of their three predictors against the actual attendance. Agents switch their selection of active predictor for the next round to be the predictor with the lowest mean prediction error over the past four rounds. This prediction, accuracy evaluation, and predictor switching process repeats each round of the simulation.

This baseline model was then modified to add a seventh predictor to the menu of available predictors for Level 0 agents. When a Level 0 agent chooses the recursive predictor, a dynamically generated set of Level 1 predictor agents is instantiated. The predictor agents represent a single-level, recursive snapshot of the current state of all the peers of the Level 0 agent that invoked them.

Each Level 1 predictor agent returns its attendance prediction based on its parent Level 0 agent's current active predictor state. The mean value of all Level 1 attendance predictions called by a Level 0 agent becomes the new attendance prediction for that Level 0 agent. In other words, a Level 0 agent that invokes the recursive predictor will compute the average of the attendance predictions of all its Level 1 predictor agents. That average attendance prediction becomes the Level 0 agent's attendance prediction for that round.

The introduction of a seventh predictor may result in less overall variance in attendance for a series of simulation runs when compared to the same model without recursive predictions. To test this, a model was configured with one hundred Level 0 agents that interacted over 50 rounds to roughly approximate a series of one year's worth of attendance predictions. That model was executed 1000 times, both with and without the recursive predictor. The time series distribution of the weekly attendance was recorded and are shown in Figure 2. The time series plots show that the system tends to converge into an adaptive ecosystem with a cyclical tendency driven by the periodic look-back nature of the specific predictors used.

As a first-order gauge of model behavior validity, the mean attendance value of 58.5 with recursion and 57.8 without recursion (see Table 1), both of which are close to the expected Nash equilibrium value of 60 predicted by Arthur [2]. The attendance distribution does exhibit clusters of data points especially at the upper end of the attendance range, which is driven by the idiosyncratic properties of the predictors selected for this model. The small number of predictors selected for this model have no inherent stochasticity and tend to produce an underlying cyclical response with sometimes sub-optimal guesses. As Garofolo

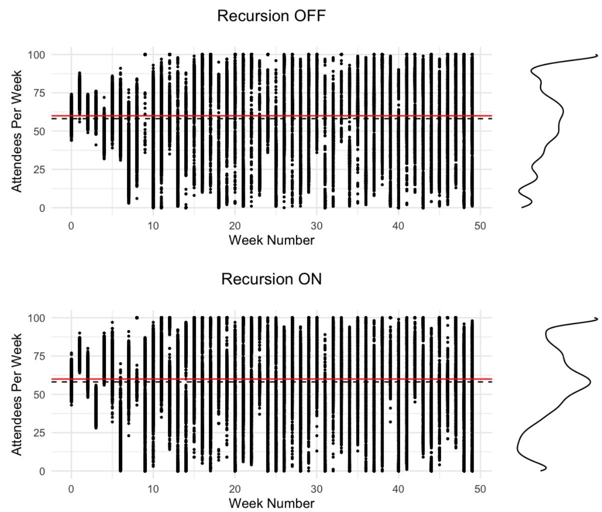

Fig. 2. Increased density around target attendance level when agents have nested ABMs compared to original model

[1999] points out, an expanded number of well-diversified predictors is important to produce smooth model performance around the Nash equilibrium point. This performance trade-off for less-than-ideal predictors was intentionally accepted in the interest of model simplicity.

The variances of the actual attendance data were compared for the same models configured with and without nesting. The model using recursive predictors showed a 5.3% reduction (Fligner-Killeen chi-squared = 131.79, df = 1, p $< .001$) in overall attendance variance compared to the configuration without recursion. These results suggest that there are conditions under which use of recursive predictors can improve the overall outcome of the model by generating predictions that have significantly less variance. Understanding more fully the regime where nesting aids performance is left for future work.

3 Recursive Agent Models L Levels Down

For an ABM representing P people, with each person considered to be a single agent, let each agent represent its environment with a subordinate ABM (sub-ABM) consisting of $P-1$ agents. The total number of ABMs involved is clearly $1+P$, while the total number of agents created is $P + P(P-1) = P^2$. Since there are P people being modeled, yet P^2 agents required, there are exactly $P^2/P = P$ copies of each individual in the overall model.

Next, consider the case of every agent in every sub-ABM having a sub-ABM running in its head, all the way down to 3 agent sub-ABMs in which each agent has a 2 agent ABM running in its head. Now the total number of ABMs in the overall model, M, is M=1+P+P(P-1)+...+P(P-1)(P-2)...3=1+eP$\Gamma(P,1) - 2\Gamma(P+1)$ where $\Gamma(k,x)$ is the incomplete gamma function, i.e.,

$$\Gamma(k,x) = \int_x^\infty t^{k+1} exp(-t)dt \tag{1}$$

and $\Gamma(k) = \Gamma(k,0)$. Similar considerations give the number of agents in these ABMs, A, as A=P+P(P-1)+P(P-1)(P-2)+...+P(P-1)(P-2)...2=eP$\Gamma(P,1) - \Gamma(P+1)$. These quantities are combinatorially large in the size of the initial population, P, placing strong upper bounds on the size of social processes that can be considered in this fashion, and limiting the number of levels that can be traversed downward, in practice. However, we emphasize that 'solving' an ABM really just means running it, so the only limitation on how large a model can be is based on available compute resources. While today this might limit us to a small number of people, surely these bounds will be relaxed as time goes on.

3.1 *ABM*s in Which Agents Have Their Own *ABM*s to Any Depth

So far the nested ABMs we have illustrated through examples have had each layer coded. Here we suggest code to automatically create all agent levels. We have experimented with a recursive formulation in which the user asks for an ABM representing P people that goes down L levels. Below is pseudo-code for such a construction. It starts by declaring Agent and ABM objects, with Agent having a link to an internal model while each Model holds an array of Agents. Declaring an ABM instance in the last line, named myModel, creates the top-level ABM via the 'ABM' constructor and then recursively generates models at each succeeding level that are one agent smaller. The execution of this code is only limited by available memory or other machine resources. As P and L increase the resources required grow exponentially, the combinatorics of these quantities implies that only a few levels are practical today on extant hardware.

```
// class definitions
forward: class ABM;
class Agent (double data;
  ABM *internalModel;
  Agent(int A, K);)
typedef Agent = *AgentPtr;
class ABM (int numberOfAgents;
  AgentPtr agents[numberOfAgents];
  ABM(int N, L);)
// constructors
Agent::Agent(int A, K) (data = 0.0;
  if (K>0) new internalModel(A, K);)
ABM::ABM(int P, L) (numberOfAgents = P;
  for (int i=0; i<P; ++i) agents[i]=new Agent(P-1,L-1))
main (ABM myModel(people, levels); return (0);)
```

We have tested a version of this code in C++ running on a large workstation, for a variety of P and L values, for an Agent object having very small size. Complete ('all the way down') models for $A = 14$ have been generated, consuming some 80 GB of RAM. For $L = 1$ we have successfully instantiated models for $A = 110$, using nearly 128 GB of RAM. Instantiating such large models takes several minutes when executed serially, and in prototype parallel (multi-threaded) implementations this has been reduced to about a minute.

4 Some Implications of Recursive *ABM*s

Formal models of social phenomena, as they have been rendered for the better part of 100 years, amount to a series of mathematical equations that need to be 'solved', in some meaningful way, to yield analytical characterizations of the kinds of phenomena under investigation. Good models closely reproduce what is known empirically about such social phenomena. An even higher bar involves models that are written in terms of the actual people who constitute the social process, i.e., models that are MI. Indeed, even in circumstances where empirical data are lacking, whether non-existent or of limited quantity or quality, the default logic in several social sciences is that models should first and foremost be methodologically individualist, and then, if data become available, they can be evaluated empirically at that time. Progress in the social sciences, given this methodology, involves coming up with agent behaviors that can be shown mathematically to yield socially-salient outcomes, at the aggregate or individual levels. A major impediment in many such investigations is that it is one thing to write down equations that are plausible behaviorally and which might produce the right results, but it can be quite another matter to solve these in anything like closed-form, or even to tease out qualitative properties from them. Analytical intractability in such situations is not merely an inconvenience, as it often is in natural science problems where resort to numerical solutions is the conventional approach in the wake of analytical difficulties, but rather is a roadblock to scientific progress as often little is known about specific functional forms or parameter values in the social sciences. In many branches of the social sciences today, analytical barriers are the main constraint limiting progress, throttling any acceleration that might be possible with the increasing availability of data and so on. While progress in mathematics may, ultimately, fuel advances in the social sciences, such developments typically arrive slowly, over decades.

However, computing with agent systems does not suffer this analytical curse. One does not 'solve' an ABM by symbolic manipulations but rather by marching a model instance forward from specific initial conditions in order to assess its results. Successful ABMs incorporate starting conditions, behavioral specifications, and parameter values that are sufficient to make the model either empirically-relevant or theoretically interesting. This 'marching forward' is not a mathematical problem but rather a computational one. Lacking the kind of barrier to progress seen in mathematical models, collections of ABMs can always be executed to reveal patterns at both the aggregate and individual levels. Only when the number of ABMs is large do problems arise in computing ‘solutions’.

As a final implication of nested agent models, consider the so-called ‘Lucas critique’ of macroeconomic models that involve policy-making agents [16]. The argument in this influential paper is that real-world economic agents have a model of policy-makers responses, and vice-versa, and this needs to be modeled to be consistent. But from our perspective this seems superficial: it is a two-level model. But perhaps in reality more layers need to be represented, as when policy-makers model investors who have their own models of the policy-makers, so 3 layers deep. One the one hand, Lucas’s critique is incomplete. On the other, his basic complaint is precisely what we are using here to argue that MI models that go down only a level or two are potentially problematic.

5 Summary and Conclusions

Methodological individualism is a hallmark of a broad branch of social science theorizing. ABMs are an emerging methodology for building social science models that are methodologically individualist [9] [11] [22], from anthropology [5], through economics [4] and finance [10], to politics [3], sociology [19] [12] and beyond. However, it is something of a curiosity that the individuals within MI models are not themselves MI. That is, social scientists who pursue methodological individualism, implicitly rejecting models that are not more aggregate in character, do not endow their agents with the same methodology. This inconsistency may seem like the only viable approach to the social science theorist who insists on deriving all results mathematically, for to perform analyses multiple levels deep, as would be required if the agents in models were themselves methodologically individualist, appears, both at first blush and upon closer inspection, to be analytically intractable. From this perspective it appears that the kind of methodological individualism on display in the economics, game theory, and finance literatures is somewhat shallow, inconsistent, even superficial, and it remains to be seen if promulgating methodological individualism down multiple levels in social science models will make a difference.

We have shown that recursive agent models can make a difference, as when unhappy agents in a Schelling segregation model run their own ABMs and make decisions on the basis of the results obtained, leading to output different from standard models. Conceptually, having only a top-level of methodological individualism is a special case of the general model involving multiple levels, since it stipulates that agents make no attempt to model other agents.

Overall, it is likely that ABMs in which agents have their own ABMs can be used to better understand certain social situations. Undoubtedly, individual humans cannot conduct actual ABM model runs in their heads. (Although future humans may use ABM models executed on computer hardware for figuring out what to do.) However, unpacking decision-making into more MI models, as when a person thinks explicitly about what other people might do/how they might behave, is almost certainly useful to consider in a wide variety of social settings.

References

1. Anderson, P.W.: More is different. Science **177**(4047), 393–396 (1972)
2. Arthur, W.B.: Inductive reasoning and bounded rationality. American Economic Review **84**(2), 406–411 (1994)
3. Axelrod, R.: The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration. Princeton University Press, Princeton, N.J. (1997)
4. Axtell, R.L.: The complexity of exchange. Economic Journal **115**, F193–210 (2005)
5. Axtell, R.L., Epstein, J.M., Dean, J.S., Gumerman, G.J., Swedlund, A.C., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., Parker, M.T.: Population growth and collapse in a multiagent model of the kayenta anasazi in long house valley. Proc Natl Acad Sci USA **99**(supplement 3), 7275–9 (2002)
6. Camerer, C.: Behavioral Game Theory. The Roundtable Series in Behavioral Economics, Princeton University Press, Princeton, N.J. (2003)
7. Coleman, J.S.: Foundations of Social Theory. The Belknap Press of Harvard University Press, Cambridge, Mass. (1990)
8. Elster, J.: Nuts and Bolts for the Social Sciences. Cambridge University Press, New York, N.Y. (1989)
9. Epstein, J.M., Axtell, R.: Growing Artificial Societies : Social Science from the Bottom Up. Brookings Institution Press/MIT Press, Washington, D.C./Cambridge, Mass. (1996)
10. Geanakoplos, J., Axtell, R.L., Farmer, J.D., Howitt, P., Conlee, B., Goldstein, J., Hendrey, M., Palmer, N.M., Yang, C.Y.: Getting at systemic risk via an agent-based model of the housing market. American Economic Review: Papers and Proceedings **102**(3), 53–58 (2012)
11. Gilbert, N.: Agent-Based Models. Quantitative Applications in the Social Sciences, Sage Publications, Inc., Thousand Oaks, CA (2008)
12. Hedstrom, P.: Dissecting the Social: On the Principles of Analytical Sociology. Cambridge University Press, New York, N.Y. (2005)
13. Kirman, A.P.: Whom or what does the representative individual represent? Journal of Economic Perspectives **6**(2), 117–136 (1992)
14. Latek, M., Kaminski, B., Axtell, R.L.: Bounded rationality via recursion (2009), cites "The Complexity of Exchange"
15. Little, D.: Varieties of Social Explanation: An Introduction to the Philosophy of Social Science. Westview Press, Summertown, Oxford, U.K. (1991)
16. Lucas, Robert E., J.: Econometric Policy Evaluation: A Critique (1976)
17. Luke, S., Balan, G.C., Panait, L.: Mason: A java multi-agent simulation library (2-4 October 2003 2003)
18. Menger, C.: Investigations into the Method of the Social Sciences. Libertarian Press, Inc., Grove City, Penn. (1985 [1883])
19. Sakoda, J.M.: The checkerboard model of social interaction. Journal of Mathematical Sociology **1**(1), 119–132 (1971)
20. Schelling, T.C.: Models of segregation. American Economic Review **59**(2), 488–493 (1969)
21. Weber, M.: Economy and Society. University of California Press, Berkeley, California (1968 [1922])
22. Wilensky, U., Rand, W.: An Introduction to Agent-Based Modeling: Modeling Natural, Social, and Engineered Complex Systems with NetLogo. MIT Press, Cambridge, Mass. (2015)