Agent Based Modelling for Social Research

Introduction to Agent Based Modelling

Welcome to the talk “Introduction to Agent-Based Modelling”, in the context of the course “Agent-based Modelling for Social Scientists”. This course is organized by the ERC-funded project “Bayesian Agent-based Population Studies”, in collaboration with the Max Planck Institute for Demographic Research.

My name is André Grow, and I am a research scientist and research area chair at the Max Planck Institute for Demographic Research. My research interests lie in digital and computational demography, with a particular focus on agent-based modelling and marriage market research.

My talk has two elements.

I will first introduce you to the basics of agent-based modelling, and I will then present an example of agent-based modelling based on my own work, to give you a better idea of how the method works and what kind of questions you can address with it.

First some definitions and basics.

As social scientists, and in particular as sociologists, we are often interested in how important macro-level social phenomena come about.

For example, scholars have been interested in understanding why metropolitan areas, such as Detroit shown here, are often racially segregated, sometimes even in the presence of laws and regulations that seek to reduce segregation.

Other scholars have been interested in understanding why there are periods in our history in which fertility was exceptionally high or low, as is the case in so-called Baby Booms and Baby Busts.
And finally, researchers have been interested in understanding why some revolutions are successful and others are not.

As diverse as these examples might appear at first glance, they have in common that they are difficult to explain, because the social systems and social mechanisms that bring them about tend to be complex.

This complexity partly derives from three factors.

First, there is interdependence, so that the decisions of one individual influence the decisions of other individuals. For example, it is by now well known that our fertility behavior is affected by the behavior of the people around us and in particular by those who are close to us, such as our friends and family. Hence, if we want to understand why a given couple has decided to have, or not to have, children, we also need to consider the fertility decisions of those around them.

Second, there often is non-linearity, which means that changes in the antecedent conditions of the social system can lead to very different outcomes. For example, modern means of communication, such as mobile phones in combination with instant messaging services today shape protests and revolutions, and the extent to which governments can control access to the internet can affect the success of protests and revolutions.

Finally, social systems often show emergent properties, that can arise from people's actions and interactions in unexpected ways and that may be difficult to square with people's stated intentions and preferences. For example, many people arguably would prefer to live in a world without human-induced climate change, but they still engage in behaviors that strongly contribute to it.

The difficulties that social complexity creates are aggravated by the fact that we typically need to consider multiple levels of society to understand how social phenomena come about.

For example, it is well known that across countries, economic development is negatively associated with overall fertility, so that fertility tends to be lower in more developed countries.

To understand how this negative association between economic development and fertility at the macro level comes about, we need to understand how at the micro level people's fertility preferences are affected by the economy, how this affects their intentions, negotiations, and actual fertility behavior within couples, and how this behavior then translates into macrolevel fertility patterns.

Studying such multi-level linkages in dynamic social systems can be difficult with traditional methods in the social science toolbox.

Agent-based modelling, by contrast, allows us to study such systems more easily, by explicitly modelling people's actions and interactions, and also modelling how these actions and interactions are affected by the macro-level contexts in which they are embedded.

Now, what is agent-based modelling and how is it defined?

Agent-based modelling is a type of computer simulation for theory development, exploration, and elaboration.

It starts from a social phenomenon, such as the negative association between economic development and fertility, and a theory that seeks to explain this phenomenon, through the actions and interactions of the people that make up the social system under consideration
A crucial and challenging aspect here is that most existing theories are formulated verbally, but not formally. This creates a problem, because agent-based modelling requires us to translate verbal theories into formal theories that can be represented in computer code. For this, we usually need to be more precise than we are when formulating a verbal theory.
For example, a verbal theory might state that improving economic conditions lead women to increasingly pursue their own careers, thereby reducing the number of children they desire.
For implementing this theory in an agent-based model, it would not be enough to simply

state that women's desired number of children decreases as the economy improves, but we would need to specify precisely how much the desired number of children decreases and how this change affects women's fertility behavior.

Once this challenging formalization step has been completed, the model is submitted to systematic simulation experiments that allow us to, first, assess the implications of the theory with analytical rigor, and, second, to create alternative what-if scenarios that enable us to assess which assumptions of the theory are crucial and what would happen if things were slightly different.

This approach also makes it possible to directly compare the explanatory power of alternative theories that rely on different assumptions. For this, we simply implement both theories in code and assess which theory is better able to explain the phenomenon of interest.

What does an agent-based model contain? What are its constituent parts?
Agent-based models usually have two main components: agents and the environment in which they are embedded.

Agents are typically autonomous, but interdependent. That is, they make their own decisions, but in doing so they are affected by the things that the other agents around them do.

Their decisions follow rules that are determined by the theory, and most models assume rather simple decisions, such as simple if-then rules. The reason why such if-then rules are often used is that already very simple rules can generate very complex social dynamics, when they are simultaneously applied by a large number of agents.
However, in some models, agent behavior is more complex. For example, agents may behave truly adaptively which means that the very nature of their decisions may change. For instance, in an economic simulation, agents might change from altruistic behavior to egoistic behavior, if their trust in others has been abused too often.

Finally, agents are embedded in an environment that can be physical, such as would be the case in a model of residential segregation, but it can also be purely social, for example when we consider how people’s fertility decisions are affected by the fertility behavior of their close friends and family. In this case, physical distances are likely less important than the precise structure of the social networks among the agents, but which of course might also be affected by physical distances.

To better illustrate what I have just described, I now present an applied example from my own research.

The example comes from the area of marriage market research, and research in this area is often interested in how the structure of the marriage market is related to observed marriage patterns.

For example, if in a given country there is a shortage of men in prime ages of marriage, maybe because of a war that cost the lives of many male soldiers, women may find it difficult to find a spouse. This difficulty, in turn, might induce them to marry men who are considerably older or younger than they are, or to forgo marriage altogether.

This is a complex social system, because we need to consider the partner preferences and partner search behavior of both men and women, and there are feedback effects, because every marriage affects the marriage market for those who are not married yet.

In this example, we are interested in the relative contributions that heterosexual men and women make to the incomes of the households that they form. When you look at these contributions, you will notice that often there is an imbalance to the disadvantage of women. What you see here on the right-hand side is for four countries the share of the household income that heterosexual women contribute to the unions that they are in. This involves both married couples and couples that live in unmarried cohabitation.

On the y-axis, you see the shares of couples that fall into different categories according to the share of the household income that the woman contributes. If on the x-axis we are close to 0, the woman has very little income and almost all of the household's income is contributed by her partner. The closer we get to 1, the more she contributes relatively to her partner to the household income. In the corners you see the share of couples in which the woman has no income at all and I excluded these numbers from the main graph to avoid that the scaling of the y-axis is too strongly affected.

When you look at these patterns, two things are notable: first, the graphs are right-skewed, so that women generally tend to earn less than their partners.

Second, there is a discontinuity at the 50/50 point. That is, the share of couples increases up to the point where she earns about as much as he does, but then there is a steep drop and there are very few couples in which she earns more than he does.

Early observers of this macro-level pattern in relative incomes have suggested that it indicates that men and women try to avoid a situation in which a woman earns more than her partner. The argument behind this is that traditional gender norms hold that the man should be the main provider for the family, so that a situation in which she earns more than he threatens the man's gender identity and therefore is avoided.

We explore an alternative mechanism that might also create such a cliff, without the need to assume that men and women are aversive of a situation where she earns more than he does. The mechanism that we propose starts from the observation that women generally earn less than men, regardless of their union status.

To illustrate this, I plot here the individual income distributions among men and women between the ages 25-45 in the same set of countries that I have shown before.

As you can see, women's individual income distributions are shifted to the left. This means that their income is on average lower than men's, but also that compared to men there are fewer women with very high incomes.

If we now assume that income is an important factor in partner selection, and also assume that both men and women prefer high-income partners over partners with lower income, these income differences imply that men and women face very different opportunities on the marriage market.

For a woman, it will be comparatively easy to find a partner who earns at least as much as she does, if not more. However, men will often have difficulties with finding partners with at least similar income, simply because there are not enough women with such incomes around.

We argue that these gender-specific differences on the marriage market can lead to the emergence of a cliff in the income distribution to the disadvantage of women, even if men and women do not evaluate a situation in which the woman earns more than the man any differently than a situation in which the man earns more.

Of course, as I told you before, marriage markets are complex dynamics systems, which makes assessing our intuition with verbal reasoning alone difficult.

For this reason, we developed a minimal marriage market model that helps us to assess the theoretical and empirical plausibility of our intuition.

The model assumes that there are 1,000 male and 1,000 female individuals, here also referred to as agents.

At the beginning of a given simulation run, these agents are single and are randomly assigned an income, based on the gender-specific income distributions observed in the 2007/2011 European Statistics on Income and Living Conditions.

After assigning agents their income, the main simulation phase starts, in which we emulate basic elements of the partner search process.

This search happens in iterations. In each iteration:

1. One male agent who is still looking for a partner is randomly paired with one female agent who is also looking for a partner.
2. During each such meeting, they need to decide whether they want to start dating each other. This is the case if they are currently dating nobody, or when the income of the person they have just met is higher than the income of their current date.
3. When both agents decide that they want to date each other, they actually start dating and leave possible other current dates for this.
4. Then, any two agents who are dating can decide that hey want to stop their search for a partner and settle for the person they are currently dating. If both decide to do so, they get married and are permanently removed from the marriage market.

Of course, this is a highly simplified representation of partner search. However, it considers the fact that partner search is a search under incomplete information and that individuals have to 'trade off' alternatives as they become available.

Importantly, we assume that income affects how likely it is that agents want to marry the person they are currently dating.

We assume that both men and women prefer partners who earn at least as much as they to over partners who earn less than they do.

This is expressed in this sigmoid function that you see here, which considers the relative income of a prospective partner (*sj*) in determining the probability that agent *i* becomes willing to marry agent *j* in a given iteration of the simulation.

M is here a constant that ensures that even individuals with no income will be able to find somebody to marry, and α is a parameter that enables us to control precisely how much male and female agents differentiate between lower or higher income partners.

This is illustrated here in this figure. In our simulations, we assume that agents strongly differentiate between partners who earn less or more than they do, which is expressed by an α-value of 30. This means that if agent *j* earns less than agent *i*, agent *i* is rather unwilling to marry *j*; yet, if *j* earns at least as much as *i*, this willingness strongly increases.

We have implemented this model in the agent-based modelling software NetLogo, and you can download the code via this tiny URL. We used it to conduct systematic simulation experiments in 27 artificial countries, each of them representing one marriage market. For each market, we repeated our simulations 50 times and averaged the outcomes across those runs, to avoid that random fluctuations affect our results.

You see here one example of these runs that uses the income distribution for Belgium. What you see in the center of the screen are the artificial men and women, respectively represented by red circles and blue triangles. At the top on the right you see the individual income distributions among men and women, and below this you see the distribution of relative incomes among the couples that have formed. This plot is now empty, as there are no couples at the beginning of a simulation run.

When I start the video, you see that gray ties form between the agents which represent the meetings that are taking place. When all agents have been paired up with an opposite sex member, they decide whether they want to start dating each other and those who start to date may even decide that they want to marry the person they have just met. If this happens, these agents are colored gray.

As the simulation proceeds, you the number of ties that form in each iteration decreases and the number of gray colored agents increases, given that more and more agents have found a partner who they have married.

At the same time, you see that the lower right graph becomes populated.

After 32 iterations, the simulation stops, because all agents have married. At this point, a strong cliff has emerged in the relative income distribution, so that in the unions that have formed, our artificial women tend to earn less or about as much as their male partners, but there are few women who earn more than their partners.

As I said before, we have applied this model to 27 different countries and you see here the average outcomes for the same four countries that I have shown before, comparing the simulation outcomes with the empirical data.

As you can see here, the results that our simulation model generates are remarkably similar to the empirical patterns, even though the model is very simple.

That is, just as we observe in the empirical data, the model generates relative income distributions across households that are right skewed. This means that there are comparatively fewer unions in which the woman earns more than the man than there are unions in which he earns more. Furthermore, and most importantly, the model generates a cliff at the 50/50 point, that often is even stronger than in the empirical data.

Thus, what our results suggest is that, given the individual income differences that exist between men and women, a gender cliff in the relative income distribution within unions can emerge, even if they do not have a preference for a situation in which the man out-earns his wife.

This has important implications for existing scholarship. Researchers have argued that the continuing increase in gender-egalitarian attitudes across the West may lead to more equality within families. However, our results suggest that even if people's partner preferences would become completely gender neutral, women may remain economically disadvantaged within their families, as long as their average income is lower than men's. Our results therefore suggest that decreasing gender inequality will not only require ideational change, but also institutional change aimed at reducing the gender pay gap, as this will create the structural conditions that are necessary to attain more equality within families.

After presenting this example, I would like to close my talk with summarizing some core points.

My first goal was to introduce you to the basics of agent-based modelling. The most important points related to this introduction are the following:

- First, social systems can be complex and agent-based modelling makes it possible to deal with this complexity.
- Second, agent-based modelling is a type of computer simulation, in which theories are explicated and assessed by means of systematic simulation experiments.
- Third, such experiments can be used for theory development and for assessing the relative merit of alternative theories, but there are many more potential purposes and uses that I cannot all cover here

My second goal was to familiarize you with an example of agent-based modelling, to illustrate how the method can be applied and what kind of questions can be answered with it.

The core points of this example were the following:

- First, at the macro level, in heterosexual unions there often is a cliff in the relative income distribution to the disadvantage of women.
- Second, this cliff does not necessarily point to traditional gender norms

Instead, agent-based modelling suggests that such a cliff can emerge from income differences between men and women, even in the absence of traditional gender norms.
With this, I have reached the end of my talk. If you have any questions of suggestions, please do not hesitate to contact me via email.