



## UvA-DARE (Digital Academic Repository)

### Socio-dynamic discrete choice: Theory and application

Dugundji, E.R.

**Publication date**  
2013

[Link to publication](#)

#### **Citation for published version (APA):**

Dugundji, E. R. (2013). *Socio-dynamic discrete choice: Theory and application*. [Thesis, fully internal, Universiteit van Amsterdam].

#### **General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

#### **Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## SUMMARY

---

Suppose you have the possibility to choose to adopt one of a number of different behaviors or to choose to buy one of a number of different products. Moreover, suppose the choice is multi-dimensional or more generally, that there are common unobserved attributes of the choice alternatives. A classic approach to statistical prediction in such a situation given an observed sample of decision making agents in a population is the nested logit model, proposed by Ben-Akiva (1973) in the context of passenger travel demand modeling and generalized by McFadden (1978) in the context of modeling the choice of residential location. The utility that a given decision making agent is presumed to associate with a particular elemental alternative is assumed to be comprised by a deterministic (to the modeler) or so-called “systematic” utility and an error term. The nested logit derives its name from the partitioning of the choice set into mutually exclusive and collectively exhaustive “nests” of elemental alternatives which are assumed to be correlated.

Now suppose your choice is additionally influenced by your individual perception of the average choices made by your neighbors, colleagues and/or socioeconomic peers. Such a specification is interesting because of the inherent dynamic that could arise if the choice model were to be applied repeatedly in successive time steps with the shares of decision makers continuously updated as a result of the choice in the previous time step. If the social influence is not important relative to other contributions in the utility, then the distribution of decision makers’ choices may not effectively change over time. However, if the social influence is strongly positive and “dominant enough” relative to other contributions in the utility, there may arise a runaway situation after several repeated time steps with a large share of decision makers flocking to one particular choice alternative. In short, the specification captures feedback between decision makers that can potentially be reinforcing over the course of time. In diverse literature this is referred to as a social multiplier, a cascade, a bandwagon effect, imitation, contagion, herd behavior, etc. (Manski 1995). Blume et al. (2011) give an extensive literature review of social interactions models in economics, including notably a section on discrete choice models of social interactions. The relative lack of attention to such models in the field of transportation is surprising. Namely, if there is theoretical or qualitative reason to believe that a feedback effect exists, it can have very important implications for the prediction of (system-wide) results over the course of time. For example, in the introduction of a new transportation mode alternative, if there is a

“dominant enough” feedback effect, this can propel the adoption of the new mode over time.

How dominant is “dominant enough” under different conditions? Our starting point in considering interdependence of various decision makers’ choices is a trio of papers by economists Aoki (1995), Brock and Durlauf (2001) and Blume and Durlauf (2003). They introduce social interactions in binary discrete choice models by allowing a given agent’s choice for a particular alternative to be dependent on the overall share of decision makers that choose that alternative. Such formulation with perfectly homogeneous agents permits the derivation of analytical results for when runaway flocking will occur. Using different approaches, all three papers show the existence of multiple equilibria depending on the values of the parameters of the model.

In moving from an idealized homogeneous agent setting towards a practical empirical setting with heterogeneous agents, there are various extensions that can be addressed, such as: i) the complexity of the discrete choice model kernel; ii) the complexity of the feedback effect; and iii) the complexity of the utility specification. Each of these extensions offers possibility to add additional types of heterogeneity to the model. In this thesis we systematically explore each of these extensions in turn.

- *Complexity of the discrete choice model kernel.* Brock and Durlauf (2002, 2006) have extended their results on the behavior of binary choice models with global social interactions to multinomial choice with global social interactions. As in the case of binary choice, also in multinomial choice, they show the existence of multiple equilibria depending on the values of the parameters of the model. In Part II of this thesis, we make Brock and Durlauf’s multinomial results precise for the case of trinary multinomial choice and extend the results for the case of nested logit with global interactions. Hereby a previously unnoticed hysteresis regime in midrange parameter space is revealed when there are more than two choice alternatives in the multinomial logit model. By considering the nested logit model, a possibility to account for unobserved heterogeneity between choice alternatives is allowed via the nesting of alternatives that are assumed to be correlated. Our analysis of the nested logit model with global social interactions yields rich bifurcation diagrams demonstrating several major additional new emergent steady state regimes where symmetry is broken by the scale parameter for the level of correlation between alternatives.
- *Complexity of the feedback effect.* While the behavior over time derived in early work assumed each decision maker to be influenced by all other decision makers (so-called global interactions), Ioannides (2006) derives more general behavior for the case of binary choice where each decision maker is influenced

by only a subset of decision makers (so-called local or non-global interactions). By considering local interactions, heterogeneity is induced since different agents will perceive different shares of agents making each choice dependent on the local network of each agent. This is important since in an empirical setting, a feedback effect defined globally would be perfectly correlated with a set of alternative specific constants in the discrete choice model. In Part III of this thesis, we illustrate the multi-agent based simulation of a discrete choice model with local interactions using microdata on transportation mode choice of households in the Netherlands as a testbed, highlighting some hypothesized network interaction effects first on the basis of abstract classes of networks in a sociodynamic binary logit model, and then on the basis of socioeconomic peer group, spatial proximity of residential location and spatial proximity of work location in a sociodynamic trinary nested logit model. Much empirical work in understanding local social-spatial feedback is still needed as the computational results prove to be highly dependent on the presumed structure of the local feedback in the econometric model estimated.

- *Complexity of the utility specification.* A key aspect in the theoretical results in Part II of this thesis is the assumption that the only observed heterogeneity in the systematic utility for the elemental choice alternatives is the feedback effect. While such a specification may be plausible for a fad, it is much less intuitive for transportation mode choice where other explanatory variables would be assumed to be significant, including attributes of the alternatives such as travel time, as well as characteristics of the decision making agents such as gender, age and income. In Part III of this thesis, a multi-agent based simulation model is therefore presented which gives straightforward possibility to test more realistic empirical cases. The multi-agent based model is docked against the analytical results in Part II for the special case of homogeneous agents as a means to verify the implementation of the computational model, before proceeding to add additional heterogeneity. In Part IV of the thesis, issues in the econometric estimation of discrete choice models with feedback effects are also explored.

We observe that the extensions of adding all three different types of heterogeneity, namely, unobserved heterogeneity captured via the nesting structure, induced heterogeneity via local interactions, and the observed heterogeneity via the fully specified systematic utility, produce dramatically different results. It can therefore be concluded undeniably that heterogeneity matters! In the remainder of this summary, we review detailed contributions.

## PART II MEAN FIELD ANALYSIS

*Sociodynamic Binary Logit: Theory*

The binary discrete choice problem with social interactions has a direct analogy in statistical physics. The choices of decision making entities can be seen as spins of atoms. With social feedback, these choices can become spontaneously polarized toward one alternative or another giving rise to observed aggregate preferences in blocks of a sampled population. In Chapter 3 we review Aoki's derivation (1995) of the binary logit model with social interactions and see that the polarization only happens however, when the parameter  $\beta$ , the coefficient characterizing the interaction effect, is higher than a certain critical value. Below this critical value, the choices are oriented at random when there are no other explanatory variables in the model, producing no net aggregate preferences. We will use these results as a benchmark later in this thesis in Chapter 6 when exploring under what circumstances the phase transition to the polarized state occurs and when it breaks down, in a case study where induced heterogeneity and observed heterogeneity are incrementally added via local interactions and via a fully specified systematic utility.

*Sociodynamic Trinary Multinomial Logit: Theory*

In Chapter 4, we apply techniques from the mathematics of dynamical systems and bifurcation theory to re-visit the multinomial logit model with social interactions originally studied by Brock and Durlauf (2002, 2006). We conclude that the sociodynamic trinary multinomial logit model has not only one, but in fact two bifurcation values of the parameter  $\beta$ , the coefficient characterizing the interaction effect. One bifurcation occurs when the Jacobian of the system has double zero eigenvalues. Another bifurcation occurs when the Jacobian has one zero eigenvalue and one negative, real eigenvalue. We find furthermore that the Jacobian of the system with  $\beta$  real, finite has no purely imaginary eigenvalues. This fact guides our search for a potential function which we derive to characterize the stability of solutions in the three regimes easily visually. We find that the regime between the two bifurcation points is an intuitively logical, but previously unnoticed hysteresis regime. The significance of this regime becomes all the more relevant when we estimate a trinary multinomial logit model with social interactions on the basis of empirical data later in this thesis in Chapter 7 and find the system in the case study indeed lies in this interesting hysteresis regime.

*Sociodynamic Trinary Nested Logit: Theory*

In Chapter 5, we study the theoretical equilibrium behavior of the trinary nested logit model with social interactions. First we observe that the trinary sociodynamic nested choice problem can be written as a two parameter planar autonomous system. By applying a graphical null clines approach, the number of solutions is counted and charted across a sweep of the utility parameter for the aggregate social feedback and a sweep of the scale parameter for the lower nest. This reveals seven qualitatively distinct solution regimes. Inspired by our success in finding a potential function in Chapter 4 for the sociodynamic multinomial logit model, next a gradient system is derived which again allows us to easily visually characterize the stability of the solutions. Five types of bifurcations are observed which describe the transitions between the regimes. Finally, the bifurcation curves which were initially observed computationally are rigorously derived analytically. This detailed theoretical work marks a starting point for a number of elaborations and variations such as higher number of choice alternatives, alternative form of choice kernel, additional observed heterogeneity, local social-spatial feedback.

The multiplicity of equilibria, and furthermore the qualitatively distinct solution regimes and the different types of bifurcations between regimes in parameter space, is clearly accentuated and becomes obvious when the only systematic part of the utility function is the feedback effect. There is thus nothing else in the utility to dampen or counter the feedback effect. Nonetheless the presence of the different regimes and their bifurcations in the “minimal” theoretical model points at underlying patterns which plausibly may exist also in fully specified empirical models with detailed systematic utility functions that account for heterogeneous choice based on individual characteristics of decision makers and individual-specific attributes of choice alternatives, and that account for local social-spatial networks – even if this is less easy to expose and detect in empirical models. It is our hope that this thesis serves as a call to other researchers, both to empiricists to try to quantify how the different solution regimes and the possible bifurcations between them affect results in empirical models, as well as to methodologists in how to further tackle the problem.

## PART III DECISION IN NETWORKS

*Sociodynamic Binary Logit: Application*

In Chapter 6, we explore a multi-agent based simulation model of binary discrete choice with interdependence of decision makers’ choices in an application to intercity travel demand using empirical data. In our approach in doing so, there is a key aspect of agent based

modeling which we believe is important to highlight as good practice, namely internal verification. Otherwise said, how can the researcher be confident that the agent based model is performing the actions that it is expected to do? What is the evidence that the programming implementation of the abstract or conceptual model is correct? To address this, we begin our modeling endeavor with the theoretical model studied in Chapter 3, building up our agent based model step by step, and deliberately adding different layers of complexity one at a time. In our case, this means adding different kinds of heterogeneity to the agent based model.

First, we consider the theoretical model where the only contribution to the systematic utility is the social influence of choices made by agents in fully connected network. We show that the emergent outcomes of this agent based model for various values of the characteristic parameter  $\beta$  yield the same results as the analytical benchmark from Chapter 3. This gives us confidence that the subtleties of scheduling, event simulation and sequences of random draws in our agent based simulation model behave as expected.

Next, we add heterogeneity to the network structure so that agents have local information, rather than global information. We experiment with two different abstract classes of networks to be able to specifically test the effect of density and of clustering. Our results with Erdős-Rényi and Watts-Strogatz graphs suggest that when a network representing the interactions between a decision making entity and the aggregate behavior of other (local) reference entities has the small world property, the system behaves in the long run as the original analytical model with global mean field information. If we would only be interested in emergent steady state outcomes and not how long the system takes to transition there, testing for the small world property may be an empirically advantageous alternative to collecting full detailed data on the precise nature of a social network.

Then, we add heterogeneity due to individual characteristics of agents (gender) as well as agent-specific attributes of choice alternatives (travel purpose, travel time, travel cost). We find that adding this additional layer of heterogeneity beyond the heterogeneity induced by the local information of the behavior of other agents in a given agent's reference group does indeed matter. The additional layer of heterogeneity dampens the social effect of the flocking behavior. The utility parameter  $\beta$  for the social influence must be significantly higher than in the minimal model in order to recover the signature phase transition.

We show furthermore that for a specific set of utility coefficients the appearance of a phase transition is robust against different instantiations of a random network at given network density, but the estimation process to determine the set of coefficients can be highly sensitive to the small variations in the different instantiations. Special

care must be taken in estimation of empirical models with networks with very low densities when the model includes self loops, and with networks with very high densities when the model includes an alternative specific constant especially in a model without self loops. In general, preference goes to models with an alternative specific constant in order to ensure the error terms in the utility function have zero mean and the estimated coefficients are unbiased. Whether self loops are implemented or not in an empirical model depends on the rationale of the system, and ideally on availability of panel data over multiple time periods.

The results presented in this chapter form the basis of further research, investigating, for example, what class of networks may be needed to address interactions between identifiable decision making entities. Technically, these may also be well modeled by interaction graphs. This is an interesting question in particular, as the results suggest that more sparse networks are more dependent on the actual reference structure. Another important direction concerns discrete choices on abstract network models with more realistic degree distributions, that is, the distribution of the number of "references" the decision making entities have.

#### *Sociodynamic Trinary Nested Logit: Application*

In Chapter 7, we continue our exploration of the multi-agent based simulation of discrete choice with interdependence of decision makers' choices, now in an application to trinary commuter transportation mode choice with various sociogeographic network scenarios. Hereby we present a framework for conceptualizing the interdependence of decision makers' choices, making a distinction between social versus spatial network interdependencies and between identifiable versus aggregate agent interdependencies. In our empirical application, we consider a model where an agent's choice is directly influenced by the proportions of the agent's neighbors, colleagues and/or socioeconomic peers making each choice. First we depict influence within a disconnected network of clustered groups. Next we depict influence within overlapping social and spatial groups. We observe that the estimated utility parameters for different hypothetical sociogeographic network scenarios can generate dramatically different dynamics. This finding underscores the need for more empirical research to understand actual sociogeographic influence networks, including those at the population level.

We also introduce observed heterogeneity in the model through sociodemographic characteristics of the agents, individual-specific attributes of the choice alternatives, and the availability of alternatives. Finally we introduce unobserved heterogeneity by accounting for common unobserved attributes of the choice alternatives in the er-

ror structure. We observe that these extensions generate dramatically different temporal dynamics and thus cannot be ignored in any true empirical application.

In order to delve deeper and disentangle the contributions of the effects, we therefore then deliberately return to our minimal nested logit model. We also take advantage of the computational possibilities permitted through social simulation of multi-agent systems to study not only the steady state behavior, but also the transition dynamics – an important needed direction for research suggested by Blume et al (2011). We study step-by-step, incrementally in turn, the effects of initial starting conditions, size effects, and the effects of various types of local interaction networks, including both clustered groups as well as overlapping groups. Doing so, we are able to draw on our earlier theoretical results to use as a benchmark in interpreting multi-agent simulation results.

With regard to effects of initial starting conditions and size effects, we demonstrate: 1) lock-in at the analytically predicted stable steady states, 2) manifestation of the analytically predicted most stable equilibrium being dominant, and 3) for a fully connected network with smaller size, the larger jump in mode share as an agent updates each choice per iteration breaks the lock-in that we found when the entire sample is fully connected.

With regard to effects of various types of local interaction networks, we demonstrate: 1) global choice behavior in a disconnected network of clustered groups is the weighted average behavior of separate clusters; 2) smaller cluster sizes yield more volatility in our nested logit model with sociodynamic feedback and as a result, a tendency towards an overall non-biased modal split averaged over many clusters; 3) with sufficient volatility, initial conditions have no significant effect on long-term outcomes; 4) with overlapping groups, influence can spread throughout entire sample; 5) for a giant cluster with sufficient network density and sufficient average degree, the precise connectivity of the network doesn't appear to matter in the long-run, but the initial conditions of the starting mode shares do matter, and the emergent distribution of outcomes for overall modal split gives the same picture as the analytically predicted outcomes for a fully connected network with the given initial conditions; 6) the analytically predicted most stable equilibrium ultimately prevails in connected, sparse network of overlapping clusters, and the initial conditions of the starting mode shares don't seem to matter in this case.

Next we study the effect of observed heterogeneity via the complexity of the utility specification on the emergent outcomes. Here we rely purely on computational techniques. It is important to recognize however that there are two stages in our process where the sociogeographic network enters. First, the network enters in the econometric estimation in determining the value of the estimated coefficients. Sec-

ond, the network enters in the multi-agent based simulation in determining the course of the spread of influence when the feedback is strong enough. We may wonder then what is the driving factor of the results: Is it simply the strength of the feedback effect relative to the other components of the utility? Or is it the connectivity of the network during the transmission process? Or both?

To gain some insight to the answer with regard to this particular case study, we run a hypothetical simulation experiment with sociogeographic networks swapped, while holding the utility parameters fixed. We find that the strength of the feedback effect relative to the other components of the utility is the dominant factor in generating the long-run results. That is, in our particular case study, the connectivity appears not to be very relevant at the transmission stage, concurring with our earlier findings in Chapter 6 for the sociodynamic binary logit application for scenarios where the network representing the interactions between a decision making entity and the aggregate behavior of other (local) reference entities has the small world property. This said, it is important to note that the sociogeographic networks studied in Chapter 7 are fairly dense by definition, due to the nature of the aggregate interaction assumed within groups. Results in Chapter 6 for a simple binary choice model with social interactions on abstract classes of networks over a sweep of network density, holding utility parameters constant, indicated that sparse networks were more sensitive in the outcomes of transmission.

#### PART IV CHALLENGES

##### *Econometric Considerations*

In Chapter 8, we discuss five strategies for introducing social and spatial network interdependencies into choice models, focusing on feedback effects and on correlated effects. Due to the nature of data available to us, we are unable to consider identifiable agent interaction in the case study. Instead we consider aggregate agent interdependencies and apply the model strategies for nine variations on three network treatments, one of these defined by socioeconomic group and two defined spatially based on residential location. According to likelihood ratio and non-nested specification tests, our best performing model strategy for this case study is what we term a “random field effect” model, namely one with unobserved individual heterogeneity on the group mean state dependence.

We believe this approach can be useful for researchers and practitioners who have a priori reason to believe there is a feedback effect in their work, for example, on theoretical or qualitative grounds. Even without such a priori knowledge, the field variable can still be a very powerful and practical means of capturing variation in a data set, es-

pecially in avoiding proliferation of estimated parameters, although strict caution must then be exercised in the interpretation of predictions over time. Access to temporal data would allow researchers and practitioners to remove ambiguity and be more definitive about their results. We hope that researchers and practitioners will start to pay more attention to the importance and consequences of agent interdependence. In particular we would be enthused to see more temporal data collected via modern electronic means such as GPS trackers and mobile phones with the aim towards understanding the potential inherent dynamic associated with feedback effects.

### *Outlook*

We have found in this thesis that careful specification of both observed and unobserved heterogeneity matters critically for emergent temporal outcomes when there is sociodynamic feedback in the model, even when the feedback takes the simple form of an aggregate field variable. Agent heterogeneity impacts the magnitude of the mode shares, the speed of the transition to the steady state as well as very fundamentally the number of possible observable steady state solutions. Also the detailed effect of induced heterogeneity is important to understand in different network structures, including the speed of information flow across them. Misrepresentation of the appropriate scale at which social influence occurs and of the appropriate network structure can yield strongly flawed policy implications when studying social feedback.

We conclude highlighting recommendations for future research, having extended previous work on discrete choice with social interactions in important ways.