Peer Effects in Random Consideration Sets

Nail Kashaev and Natalia Lazzati*†

May 23, 2019

(First Version April 13 arXiv:1904.06742)

Abstract

This paper develops a dynamic model of discrete choice that incorporates peer effects into consideration sets. We characterize the equilibrium behavior and study the empirical content of the dynamic model we offer. In our set-up, the choices of friends act as exclusion restrictions. They provide the variation in the consideration sets that we exploit to recover the ranking of preferences of each person, the attention mechanism, and the set of connections between the people in the network. The identification strategy we offer does not rely on the variation of the set of available options (or menus), which remain the same across all the observations.

JEL codes: D83, O33

Keywords: Peer Effects, Consideration Sets, Continuous Time Markov Process

*Nail Kashaev, nkashaev@uwo.ca, University of Western Ontario; Natalia Lazzati, UC Santa Cruz, *e-mail:* nlazzati@ucsc.edu.

†The authors thank Victor Aguiar for his useful comments.

1 Introduction

Much research in economics has found that peer effects have an important role in explaining people′s choices. The basic idea is that a person is more likely to select a specific option if more of her friends are doing so. Models of peer effects typically assume that the person is aware of all the available options and the choices of friends affect her preference ranking. The more recent literature on (single-agent) consideration set models relaxes this full awareness postulate by allowing people to consider only a subset of the available options when making a decision. We develop a dynamic model of discrete choice that incorporates peer effects in the formation of consideration sets. In doing so, we provide an alternative mechanism for peer effects. The model we offer is quite tractable from an applied perspective. In particular, we show that all its primitives can be uniquely recovered from a sequence of choices. These primitives include the strict preference ranking of each person, the attention mechanism, and (surprisingly) the set of connections between the people in the network!

In our set-up, people are linked through a social network that captures the interactions between individuals. Each person in the network has a strict preference order over a finite set of options or alternatives. At random moments, a given person can revise her current choice and pick up a new option. The person sticks to this option till the revision opportunity arises again. We assume the person is boundedly rational and does not consider all the available options at the moment of revising her selection. Instead, she observes the choices of her friends and forms a consideration set. Then the person selects the most preferred option among the ones she is actually considering. This model leads to a sequence of choices that evolve through time according to a continuous-time Markov process.

We initially show the dynamic system has a unique equilibrium (or invariant distribution). We do so under the assumption that each option has (a priori) nonzero probability of being considered by each person irrespective of the choices of her peers. This assumption captures the idea that a person can eventually learn about an alternative in many different ways (outside the control of our model) including the possibility of watching an ad on television.

It assures that we can move from any initial configuration of choices to any other one in finite time. We then show that the model primitives are uniquely identified up to conditional choice probabilities of agents. In the single-agent consideration set models these conditional choice probabilities are usually directly observed objects. Since we deal with a continuous-time dynamic model, we need to take extra steps and establish identification of the conditional choice probabilities as well. We do so for alternative datasets.

We build the identification results in a sequence of steps. First, we assume that the conditional choice probabilities can be recovered from the data. Each of these probabilities informs us about the frequency of choices of a given person conditional on the alternatives selected by others. In our model, a person is more likely to pay attention to a specific option if more of her friends are currently adopting it. Thus, the choices of friends act as exclusion restrictions. They provide the variation in the consideration sets that we exploit to recover the set of connections between the people in the network and their ranking of preferences. We can then use this information to recover the attention mechanism of each person, i.e., the probability of including a specific option in the consideration set as a function of the number of friends who are currently choosing it. Interestingly, the identification strategy we pursue does not rely (as most of the theoretical work on consideration sets) on variation of the set of available options (or menus).

Second, we study identification of the conditional choice probabilities. We consider three datasets that differ regarding its informational content: continuous-time data; discrete-time data with arbitrary time intervals; and the distribution of equilibrium choices. The first two datasets allow us to recover the transition rate matrix of the dynamic system (also known as the infinitesimal generator matrix in the statistical literature), and from there we can identify the conditional choice probabilities. In the case of continuous-time data the transition rate matrix is identified without any extra restrictions. To identify the transition rate matrix using discrete-time data with arbitrary time intervals, we invoke insights from Blevins (2017, 2018). The driven force of this result is that the transition rate matrix in our model is rather parsimonious. In particular, the selection revision process we use is such that the probability that two or more people revise their selected options at the same time is zero. This property translates into a transition rate matrix that has zeros in many known locations.

In the third dataset, we study the possibility of recovering the conditional choice probabilities from equilibrium behavior. (This dataset is clearly less informative than the two previous ones.) To this end, we first show that, under symmetry restrictions, the equilibrium behavior of our model coincides with the so-called Gibbs equilibrium. In this context, identification follows immediately. In particular, the conditional choice probabilities of each person coincide with the corresponding conditional probabilities obtained from equilibrium behavior. We offer some insights to extend this idea to the heterogeneous case.

All previous results rely on deterministic preferences. We then show these results extend to the case of stochastic preferences (in addition to stochastic consideration sets) if each person has a few connections. We finally illustrate the main ideas using a simple model of restaurant choice. This illustration highlights the role of the network structure in shaping people mistakes, e.g., it shows that homophyly reduces the frequency of mistakes. It also highlights some estimation aspects of our model.

From a modelling perspective, our set-up combines the dynamic model of social interactions of Blume (1993, 1995) with the (single-agent) model of random consideration sets of Manzini and Mariotti (2014). By adding peer effects into the consideration sets we are able to use the choices of others as instruments to recover preferences. As we mentioned above, the literature on identification of single-agent consideration set models has mainly relied on variation of the set of available options or menus. The latter includes Aguiar (2017), Aguiar et al. (2016), Brady and Rehbeck (2016), Caplin et al. (2018), Cattaneo et al. (2017), Horan (2018), Lleras et al. (2017), Manzini and Mariotti (2014), and Masatioglu et al. (2012).[1] (See Aguiar et. al (2019) for a comparison of several consideration set models in an ex-

[1] See also Manski (1977) for a throughout formulation of the discrete choice model that incorporates the possibility that the decision maker only considers a sub-set of options.

perimental setting.) Other papers have relied on the existence of exogenous covariates that shift preferences or consideration sets. The latter include Barseghyan et al. (2019), Conlon and Mortimer (2013), Dranganska and Klapper (2011), Gaynor et al. (2016), Goeree (2008), Mehta et al. (2003), and Roberts and Lattin (1991). Variation of exogenous covariates has also been used by Abaluck and Adams (2017) via an approach that exploits symmetry breaks with respect to the full consideration set model.

As we also mentioned, we can recover from the data the set of connections between the people in the network. In the context of linear models, a few recent papers have made progress in the same direction. Among them, Blume et al. (2015), Bonaldi et al. (2015), De Paula et al. (2018), and Manresa (2013). In the context of discrete-choice, Chambers et al. (2019) also identify the network structure but in their model peer effects do not affect consideration sets but preferences (among other differences).

The connection between the equilibrium behavior in our model and the Gibbs equilibrium is similar to the one in Blume and Durlauf (2003).

Let us finally mention two other papers that incorporate peer effects in the formation of consideration sets: Borah and Kops (2018) do so in a static framework and rely on variation of menus for identification. Lazzati (2018) considers a dynamic model but the time is discrete and she focuses on two binary options that can be acquired together.

The rest of the paper is organized as follows. Section 2 presents the model and describes the equilibrium behavior. Section 3 studies the empirical content of the model. Section 4 extends the initial idea to contemplate random preferences in addition to random consideration sets. Section 5 presents some simulation results for a model of choosing a restaurant. Section 6 concludes, and all the proofs are collected in Section 7.

2 The Model

2.1 Social Network, Consideration Sets, and Choices

There is a finite set of people connected through a social network. The network is described by a simple graph $\Gamma = (\mathcal{A}, e)$, where $\mathcal{A} = \{1, 2, ..., A\}$ is the set of nodes (or people) and e is the set of edges. Each edge identifies two connected people and the direction of the connection. For each person $a \in \mathcal{A}$, her set of friends (or reference group) is defined as follows

$$\mathcal{N}_a = \{a' \in \mathcal{A} : a' \neq a \text{ and } a' \text{ is connected to } a \text{ through an edge in } \Gamma\}.$$

There is a set of alternatives $\overline{\mathcal{Y}} = \mathcal{Y} \cup \{o\}$ from which each person might choose, where $\mathcal{Y} = \{1, 2, ..., Y\}$ is a finite set of options and o is a default option. Each person a has a strict preference order $\succ_a$ over the set of options $\mathcal{Y}$. All people agree in that the default option is the least preferred. We refer to $\mathbf{y} = (y_a)_{a \in \mathcal{A}} \in \overline{\mathcal{Y}}^A$ as a choice configuration.

We model the revision process of alternatives as a standard continuous-time Markov process on the space of choice configurations that describes the evolution of people choices through time. In particular, we assume that people are endowed with independent Poisson "alarm clocks" with rates $\boldsymbol{\lambda} = (\lambda_a)_{a \in \mathcal{A}}$. At randomly chosen moments (exponentially distributed with mean $1/\lambda_a$) the alarm of person a goes off.[2][3] When this happens, the person selects the most preferred alternative among the ones she is actually considering. Let us indicate by $\mathcal{C} \subseteq \mathcal{A}$ her consideration set. The decision rule of person a can be summarized by a simple indicator function

$$\mathrm{R}_a(v|\,\mathcal{C}) = 1(v \succ_a v' \text{ for all } v' \in \mathcal{C})$$

that takes value 1 if v is the most preferred alternative in $\mathcal{C}$ according to $\succ_a$. If, at the

[2] That is, each person a is endowed with a collection of random variables $\{\tau_n^a\}_{n=1}^{\infty}$ such that each difference $\tau_n^a - \tau_{n-1}^a$ is exponentially distributed with mean $1/\lambda_a$. All these differences are independent across people and time.

[3] See Blume (1993, 1995) for theoretical models that rely on Poisson "alarm clocks" and Blevins (2018) for a nice discussion of the advantages of this type of revision process from an applied perspective.

moment of choosing, the consideration set of person a does not include any alternative in $\mathcal{Y}$, then the person selects the default option.

In our model, whether person a pays attention to a particular alternative depends on the configuration of choices of her friends at the moment of revising her selection. Let $\mathrm{N}_a^v(\mathbf{y})$ be the number of friends of person a who select option v in choice configuration $\mathbf{y}$. Formally,

$$\mathrm{N}_a^v(\mathbf{y}) = \sum\nolimits_{a' \in \mathcal{N}_a} 1(y_{a'} = v).$$

The probability that person a pays attention to alternative $v \in \mathcal{Y}$ given a choice configuration $\mathbf{y}$ is $\mathrm{Q}_a(v|\mathrm{N}_a^v(\mathbf{y}))$. That is, whether the person pays attention to a specific alternative depends on how popular that alternative is among her group of reference. It follows that the probability of facing consideration set $\mathcal{C}$ is

$$\prod\nolimits_{v \in \mathcal{C}} \mathrm{Q}_a(v|\mathrm{N}_a^v(\mathbf{y})) \prod\nolimits_{v \notin \mathcal{C}} (1 - \mathrm{Q}_a(v|\mathrm{N}_a^v(\mathbf{y}))).$$

By combining preferences and stochastic consideration sets, the (ex-ante) probability that person a selects (at the moment of choosing) alternative $v \in \mathcal{Y}$ is given by

$$\mathrm{P}_a(v|\mathbf{y}) = \sum\nolimits_{\mathcal{C} \subseteq 2^{\mathcal{Y}}} \mathrm{R}_a(v|\,\mathcal{C}) \prod\nolimits_{v' \in \mathcal{C}} \mathrm{Q}_a\left(v'|\mathrm{N}_a^{v'}(\mathbf{y})\right) \prod\nolimits_{v' \notin \mathcal{C}} \left(1 - \mathrm{Q}_a\left(v'|\mathrm{N}_a^{v'}(\mathbf{y})\right)\right).$$

When preferences are deterministic, this last expression simplifies to

$$\mathrm{P}_a(v|\mathbf{y}) = \mathrm{Q}_a(v|\mathrm{N}_a^v(\mathbf{y})) \prod\nolimits_{v' \in \mathcal{Y}, v' \succ_a v} \left(1 - \mathrm{Q}_a\left(v'|\mathrm{N}_a^{v'}(\mathbf{y})\right)\right). \tag{1}$$

The probability of selecting the default option o is just $\prod_{v \in \mathcal{Y}} (1 - \mathrm{Q}_a(v|\mathrm{N}_a^v(\mathbf{y})))$. That is, the default option is selected only when the consideration set is empty.

Let us add a few comments about our model. First, it represents truly boundedly rational agents. The people in our framework do not solve a dynamic optimization problem and their choice sets may thereby not include their most preferred alternatives for long periods of time. Second, in our initial specification, the only source of randomness in choice is via consideration sets. In this sense, our initial model captures a single, though important, channel of possible mistakes in choices. The social network shapes the nature and the strength of these mistakes.

We extend the analysis to random preferences in Section 4. In this extension, $\mathrm{R}_a(\cdot| \mathcal{C})$ is not an indicator function but a distribution on $\mathcal{Y}$.

Remark. Our attention mechanism assumes that the probability of paying attention to a given option depends on the choices of peers at the moment of revising the selection. It is independent of the current selection of the person. The model can be modified to allow the current choice of the agent to affect her next choice when the revision moment arrives. All our results would go through with mild modifications except the connection of our model with the Gibbs random field models.

2.2 Equilibrium

The independent identically distributed (i.i.d.) Poisson "alarm clocks", which lead the selection revision process, guarantee that at each time interval at most one person revises her selection almost surely. Thus, the transition rates between choice configurations that differ in more than one component are zero. Formally, the transition rate from choice configuration $\mathbf{y}$ to any different one $\mathbf{y}'$ is as follows

$$\mathrm{m}(\mathbf{y}' \mid \mathbf{y}) = \begin{cases} 0 & \text{if } \sum_{a\in\mathcal{A}} 1(y'_a \neq y_a) > 1 \\ \sum_{a\in\mathcal{A}} \lambda_a \mathrm{P}_a(y'_a|\mathbf{y}) 1(y'_a \neq y_a) & \text{if } \sum_{a\in\mathcal{A}} 1(y'_a \neq y_a) = 1 \end{cases}. \tag{2}$$

In the statistical literature on continuous-time Markov processes these transition rates are the out of diagonal terms of the *transition rate matrix* (also known as the *infinitesimal generator matrix*). The rate of transition out from a given choice configuration $\mathbf{y}$ is simply

$$\mathrm{m}(\mathbf{y} \mid \mathbf{y}) = -\sum\nolimits_{\mathbf{y}'\in\overline{\mathcal{Y}}^A\setminus\{\mathbf{y}\}} \mathrm{m}(\mathbf{y}' \mid \mathbf{y}).$$

We will indicate by $\mathcal{M}$ the transition rate matrix. In our model, the number of choice configurations is $(Y+1)^A$. Thus, $\mathcal{M}$ is a $(Y+1)^A \times (Y+1)^A$ matrix. There are many different ways of ordering the choice configurations and thereby writing the transition rate matrix. To avoid any sort of ambiguity in the exposition, we will let the choice configurations be ordered according to the lexicographic order with o treated as zero. Constructed in this way the first element of $\mathcal{M}$ is (for instance) $\mathcal{M}_{11} = \mathrm{m}\big((o, o, ..., o)' \mid (o, o, ..., o)'\big)$. Formally, let $\iota(\mathbf{y}) \in \left\{1, 2, ..., (Y+1)^{A}\right\}$ be the position of $\mathbf{y}$ according to the lexicographic order. Then,

$$\mathcal{M}_{\iota(\mathbf{y})\iota(\mathbf{y}')} = \mathrm{m}\left(\mathbf{y}' \mid \mathbf{y}\right).$$

An equilibrium in our model is an invariant distribution $\mu : \overline{\mathcal{Y}}^{A} \to [0,1]$, with $\sum_{\mathbf{y}\in\overline{\mathcal{Y}}^{A}} \mu(\mathbf{y}) = 1$, of the dynamic process with transition rate matrix $\mathcal{M}$. It indicates the likelihood of each choice configuration $\mathbf{y}$ in the long run. This equilibrium behavior relates to the transition rate matrix in a linear fashion

$$\mu\mathcal{M} = \mathbf{0}.$$

To guarantee existence of such an equilibrium we impose a simple restriction. This extra assumption will also play a key role in the identification of the model.

(A1) For each $a \in \mathcal{A}$, $v \in \mathcal{Y}$, and $\mathbf{y} \in \overline{\mathcal{Y}}^{A}$,

$$1 > \mathrm{Q}_{a}\left(v|\mathrm{N}_{a}^{v}(\mathbf{y})\right) > 0.$$

Assumption A1 simply states that, for any choice configuration, the probability that any given option ends up in the consideration set of each person at the moment of revising her selection is strictly positive. This assumption captures the idea that a person can eventually learn about an alternative in many different ways (outside the control of our model) including the possibility of watching an ad on television. It follows from A1 that each subset of options is (ex-ante) considered with nonzero probability.

Below, we let $\mu_{-a}(\mathbf{y}_{-a}) = \sum_{v\in\overline{\mathcal{Y}}} \mu(v, \mathbf{y}_{-a})$ with $\mathbf{y}_{-a} = (y_{a'})_{a'\in\mathcal{A}\setminus\{a\}} \in \overline{\mathcal{Y}}^{A-1}$. Proposition 1 states equilibrium existence and characterizes equilibrium behavior.

Proposition 1 *If A1 is satisfied, then there exists a unique μ. Also, μ satisfies*

$$\mu(\mathbf{y}) = \frac{1}{\sum_{a\in\mathcal{A}}\lambda_{a}} \sum_{a\in\mathcal{A}} \lambda_{a} P_{a}(y_{a}|\mathbf{y})\, \mu_{-a}(\mathbf{y}_{-a}) \;\; \textit{for each } \mathbf{y} \in \overline{\mathcal{Y}}^{A}.$$

The next example describes the equilibrium behavior of a simple specification of our model.

Example 1: There are two identical, connected people that select among two alternatives, namely, option 1 and the default option o. The rates for their Poisson "alarm clocks" are 1. Thus, for $a = 1, 2$, we get that

$$\mathrm{P}_a\left(1|\mathbf{y}\right) = \mathrm{Q}\left(1|\mathrm{N}_a^v\left(\mathbf{y}\right)\right) \text{ and } \mathrm{P}_a\left(o|\mathbf{y}\right) = 1 - \mathrm{Q}\left(1|\mathrm{N}_a^v\left(\mathbf{y}\right)\right).$$

Note that we avoided the sub-index in Q because of the symmetry.

The transition rate matrix $\mathcal{M}$ is as follows. (The columns are ordered as the rows.)

(o, o)	$-2\mathrm{Q}\left(1\|0\right)$	$\mathrm{Q}\left(1\|0\right)$	$\mathrm{Q}\left(1\|0\right)$	0
$(o, 1)$	$1 - \mathrm{Q}\left(1\|0\right)$	0	$-1 + \mathrm{Q}\left(1\|0\right) - \mathrm{Q}\left(1\|1\right)$	$\mathrm{Q}\left(1\|1\right)$
$(1, o)$	$1 - \mathrm{Q}\left(1\|0\right)$	$-1 + \mathrm{Q}\left(1\|0\right) - \mathrm{Q}\left(1\|1\right)$	0	$\mathrm{Q}\left(1\|1\right)$
$(1, 1)$	0	$1 - \mathrm{Q}\left(1\|1\right)$	$1 - \mathrm{Q}\left(1\|1\right)$	$-2 + 2\mathrm{Q}\left(1\|1\right)$

.

After simple calculations, the steady-state equilibrium is given by

$$\begin{aligned} \mu\left(o, o\right) &= \frac{\left[1 - \mathrm{Q}\left(1|0\right)\right]\left[1 - \mathrm{Q}\left(1|1\right)\right]}{\blacktriangle} \\ \mu\left(o, 1\right) = \mu\left(1, o\right) &= \frac{\mathrm{Q}\left(1|0\right)\left[1 - \mathrm{Q}\left(1|1\right)\right]}{\blacktriangle} \\ \mu\left(1, 1\right) &= \frac{\mathrm{Q}\left(1|0\right)\mathrm{Q}\left(1|1\right)}{\blacktriangle} \end{aligned}$$

with $\blacktriangle = 1 - \mathrm{Q}(1|1) + \mathrm{Q}(1|0)$. ■

3 Empirical Content of the Model

This section provides conditions under which the researcher can uniquely recover (from the data) the set of connections $\Gamma = (\mathcal{A}, e)$, the profile of strict preferences $(\succ_a)_{a\in\mathcal{A}}$, the attention mechanism $(\mathrm{Q}_a)_{a\in\mathcal{A}}$, and the rates of the Poisson "alarm clocks" $(\lambda_a)_{a\in\mathcal{A}}$. We offer alternative conditions under which the model is identified. The requirements we propose vary with the strength of the datasets we consider. As it is always the case with identification, we will abstract from small sample issues.

We will separate the identification analysis in two parts. First, we will assume the researcher knows the conditional choice probabilities $(\mathrm{P}_a)_{a\in\mathcal{A}}$ and will provide conditions under which the main parts of the model can be uniquely recovered from this information. We will then elaborate on the identification of the conditional choice probabilities $(\mathrm{P}_a)_{a\in\mathcal{A}}$.

3.1 Identification of the Model Knowing $(\mathbf{P}_a)_{a\in\mathcal{A}}$

Let us initially assume the researcher knows the conditional choice probabilities $(\mathrm{P}_a)_{a\in\mathcal{A}}$. Our identification strategy relies on two extra assumptions.

(A2) For each $a \in \mathcal{A}$, $|\mathcal{N}_a| > 0$.

(A3) For each $a \in \mathcal{A}$ and $v \in \mathcal{Y}$, $\mathrm{Q}_a(v|k)$ is strictly increasing in k.

Assumption A2 requires each person to have at least one friend. Assumption A3 states that each person pays more attention to a particular option if more of her friends are adopting it. Under assumptions A1-A3, the choices of peers act as exclusion restrictions in the stochastic variation of the consideration sets. This variation allows us to recover the set of connections between the people in the network and the preference ranking of each of them. We can then sequentially identify the attention mechanism of each person moving from the most preferred alternative to the least preferred one. Proposition 2 formalizes these claims.

Proposition 2 *Suppose A1-A3 are satisfied and we know* $(P_a)_{a\in\mathcal{A}}$. *Then, the set of connections* $\Gamma = (\mathcal{A}, e)$*, the profile of strict preferences* $(\succ_a)_{a\in\mathcal{A}}$*, and the attention mechanism* $(Q_a)_{a\in\mathcal{A}}$ *are identified.*

The next example sheds extra light on the identification procedure we use in the last proposition.

Example 2: Suppose there are three people $\mathcal{A} = \{1,2,3\}$ that select among two alternatives $\mathcal{Y} = \{1,2\}$ and the default option o. The researcher knows P_1, P_2, and P_3. Let us consider person 1. The probability that Person 1 selects the default option o (given a profile of choices $\mathbf{y}$) is

$$\mathrm{P}_1(o|\mathbf{y}) = \left(1 - \mathrm{Q}_1\left(1|\mathrm{N}_1^1(\mathbf{y})\right)\right)\left(1 - \mathrm{Q}_1\left(2|\mathrm{N}_1^2(\mathbf{y})\right)\right).$$

Under A3, we get that $2 \in \mathcal{N}_1$ if and only if

$$\mathrm{P}_1(o|o,o,o) > \mathrm{P}_1(o|o,1,o).$$

Similarly, $3 \in \mathcal{N}_1$ if and only if $\mathrm{P}_1(o|o,o,o) > \mathrm{P}_1(o|o,o,1)$. Thus, we can learn from the data the set of friends of Person 1. Let us assume we get that $\mathcal{N}_1 = \{2\}$. To recover the preferences of Person 1 note that

$$\begin{array}{ll} \mathrm{P}_1(1|\mathbf{y}) = \mathrm{Q}_1\left(1|\mathrm{N}_1^1(\mathbf{y})\right) & \text{if} \quad 1 \succ_1 2 \\ \mathrm{P}_1(1|\mathbf{y}) = \mathrm{Q}_1\left(1|\mathrm{N}_1^1(\mathbf{y})\right)\left(1 - \mathrm{Q}_1\left(2|\mathrm{N}_1^2(\mathbf{y})\right)\right) & \text{if} \quad 2 \succ_1 1 \end{array}.$$

Thus, $2 \succ_1 1$ if and only if

$$\mathrm{P}_1(1|o,o,o) > \mathrm{P}_1(1|o,1,o).$$

Suppose that, indeed, we get that $2 \succ_1 1$. We can finally recover the attention mechanism via the next four probabilities in the data

$$\begin{array}{ll} \mathrm{P}_1(2|o,o,o) = \mathrm{Q}_1(2|0) & \mathrm{P}_1(2|o,2,o) = \mathrm{Q}_1(2|1) \\ \mathrm{P}_1(1|o,o,o) = \mathrm{Q}_1(1|0)(1 - \mathrm{Q}_1(2|0)) & \mathrm{P}_1(1|o,1,o) = \mathrm{Q}_1(1|0)(1 - \mathrm{Q}_1(2|1)) \end{array}.$$

By a similar exercise we can recover the relevant information of Persons 2 and 3. ■

3.2 Identification of $(\mathbf{P}_a)_{a \in \mathcal{A}}$

This section studies identification of the conditional choice probabilities and the rates of the Poisson "alarm clocks" from three different datasets.

First we assume researcher observes people′s choices at time intervals of length Δ and can consistently estimate $\Pr\left(\mathbf{y}^{t+\Delta} = \mathbf{y}' \mid \mathbf{y}^t = \mathbf{y}\right)$ for each pair $\mathbf{y}', \mathbf{y} \in \overline{\mathcal{Y}}^A$. We will capture these transition probabilities by a matrix $\mathcal{P}(\Delta)$. (Here again, we will assume that the choice configurations are ordered according to the lexicographic order when we construct $\mathcal{P}(\Delta)$.) The connection between $\mathcal{P}(\Delta)$ and $\mathcal{M}$ is

$$\mathcal{P}(\Delta) = e^{(\Delta \mathcal{M})}.$$

The first two datasets we consider only differ regarding Δ. Specifically, in the first dataset we let the time interval be very small. We can think of this dataset as the "ideal dataset" that registers people′s choices in continuous time. With the proliferation of on-line platforms and scanner this sort of data might indeed be available for some applications! In the second dataset we allow the time interval to be of arbitrary size. In the third dataset we assume the researcher can only recover the distribution of equilibrium choices. The informational content clearly decreases as we move from the first to the last dataset.

The next table formally describes the three datasets we consider.

Dataset 1	The researcher knows $\lim_{\Delta \to 0} \mathcal{P}(\Delta)$
Dataset 2	The researcher knows $\mathcal{P}(\Delta)$
Dataset 3	The researcher knows μ

The first result of this section is as follows.

Proposition 3 (Dataset 1) *The conditional choice probabilities* $(P_a)_{a \in \mathcal{A}}$ *and the rates of the Poisson "alarm clocks"* $(\lambda_a)_{a \in \mathcal{A}}$ *are identified.*

Remark. The proof of Proposition 3 relies on the fact that when the time interval between the observations goes to zero, then we can recover $\mathcal{M}$. There are at least two well-known cases that produce the same outcome without requiring $\Delta \to 0$. The first one happens when the length interval Δ is below a threshold $\overline{\Delta}$. The second one occurs when the researcher can observe the dynamic system at two different intervals Δ_1 and Δ_2 that are not multiple of each other. (See, e.g., Blevins (2017) and the literature therein.)

The next proposition states that, by adding an extra restriction, the transition rate matrix can be identified from people′s choices even if these choices are observed at the endpoints of discrete time intervals. In this case, the researcher needs to know the rates of the Poisson "alarm clocks", or normalize them in empirical work.

Proposition 4 (Dataset 2) *If A2 is satisfied, the researcher knows $(\lambda_a)_{a\in\mathcal{A}}$ and $\mathcal{M}$ has distinct eigenvalues that do not differ by an integer multiple of $2\pi i/\Delta$, then the conditional choice probabilities $(P_a)_{a\in\mathcal{A}}$ are generically identified.*

The key element in proving Proposition 4 is that the transition rate matrix of our model is rather parsimonious. To see why, recall that, at any given time, only one person revises her selection with nonzero probability. This feature of the model translates into a transition rate matrix $\mathcal{M}$ that has many zeros in known locations.

We finally discuss identification of the conditional choice probabilities from Dataset 3. This discussion relies on the connection between our results with the Gibbs random field models. These models have been used to study social interactions by Allen (1982), Blume (1993, 1995), and Blume and Durlauf (2003), among many others. We will use the connection between the two models to discuss the identification of $(\mathrm{P}_a)_{a\in\mathcal{A}}$ from equilibrium behavior μ. To this end we will assume the rates of the Poisson "alarm clocks" are identical for all people. In this case, by Proposition 1, the equilibrium behavior μ relates linearly to the conditional choice probabilities $(\mathrm{P}_a)_{a\in\mathcal{A}}$

$$\mu(\mathbf{y}) = \frac{1}{A}\sum\nolimits_{a\in\mathcal{A}} \mathrm{P}_a(y_a|\mathbf{y})\,\mu(\mathbf{y}_{-a}) \text{ for each } \mathbf{y} \in \overline{\mathcal{Y}}^A.$$

The main difficulty for identification (if we only observe μ) is that the number of moments in the data is usually smaller than the number of expressions we want to recover. This can be easily seen if we eliminate the symmetry condition in Example 1. In this case, we would be interested in recovering four conditional choice probabilities

$$\mathrm{P}_1(1|0),\ \mathrm{P}_1(1|1),\ \mathrm{P}_2(1|0),\ \text{and } \mathrm{P}_2(1|1).$$

(The conditional probabilities of choosing the default option can be obtained directly from the latter.) In this illustration, the dataset would contain four equilibrium moments, namely, $\mu(o,o)$, $\mu(1,o)$, $\mu(o,1)$, and $\mu(1,1)$. But only three of them can be linearly independent. So, in this case, the choice probabilities we are interested in would be just partially identified. This issue can be solved if we add symmetry across people. If we do so, we would have only two elements to recover, and the model would be overidentified. Moreover, the conditional choice probabilities would relate to the equilibrium conditions in a simple way

$$\mathrm{P}(1|0) = \mu(1,o) / [\mu(o,o) + \mu(1,o)] \text{ and } \mathrm{P}(1|1) = \mu(1,1) / [\mu(1,1) + \mu(o,1)].$$

That is, the conditional choice distributions of each person coincide with the corresponding conditional distributions obtained from equilibrium behavior. This result can be extended to the case of more options and/or more people by using the notion of compatibility of conditional distributions that we include next for completeness.

Definition: We say $(\mathrm{P}_a)_{a\in A}$ is a set of compatible conditional distributions if there exists a joint distribution P: $\overline{\mathcal{Y}}^A \to [0,1]$, with $\sum_{\mathbf{y}\in\overline{\mathcal{Y}}^A}\mathrm{P}(\mathbf{y}) = 1$, such that

$$\mathrm{P}_a(y_a|\mathbf{y}) = \mathrm{P}(\mathbf{y}) / \sum\nolimits_{y_a\in\overline{\mathcal{Y}}} \mathrm{P}(\mathbf{y}) \text{ for each } \mathbf{y} \in \overline{\mathcal{Y}}^A.$$

The last identification result follows from connecting the equilibrium behavior in our model with the Gibbs equilibrium. A similar connection is discussed in Blume and Durlauf (2003).

Proposition 5 (Dataset 3) *If $(P_a)_{a\in A}$ is a set of compatible conditional distributions and the rates of the Poisson "alarm clocks" are identical for all people, then the conditional choice probabilities $(P_a)_{a\in\mathcal{A}}$ are identified. Moreover,*

$$P_a(y_a|\mathbf{y}) = \mu(\mathbf{y}) / \mu_{-a}(\mathbf{y}_{-a}) \textit{ for each } \mathbf{y} \in \overline{\mathcal{Y}}^A.$$

The technical conditions required for a set of conditional distributions to be compatible are discussed in Kaiser and Cressie (2000). Their analysis implies that compatibility demands strong symmetric restrictions. In particular, in the two people, two actions case, Arnold and Press (1989) show that compatibility holds if and only if the next equality is satisfied

$$\frac{1-\mathrm{Q}_1(1|0)}{\mathrm{Q}_1(1|0)}\frac{\mathrm{Q}_1(1|1)}{1-\mathrm{Q}_1(1|1)}=\frac{1-\mathrm{Q}_2(1|0)}{\mathrm{Q}_2(1|0)}\frac{\mathrm{Q}_2(1|1)}{1-\mathrm{Q}_2(1|1)}.$$

Thus, while the identification strategy in Proposition 5 is interesting in that it only requires data on equilibrium behavior, its drawback is the symmetry restrictions that might not be quite appealing in practice. We next explain that (in our setting) these restrictions could be relaxed if the network is complete and large.

Assume that all people are connected. The number of people in the network is A. For each of them, there are Y available options (in addition to the default one). Thus, the number of equilibrium points minus one (since all of the probabilities add up one) is

$$(Y+1)^A-1.$$

This number imposes an upper bound on the number of conditional choice probabilities we can recover from the data. Recall that, for each alternative $v\in\mathcal{Y}$,

$$\mathrm{P}_a(v|\mathbf{y})=\mathrm{Q}_a(v|\mathrm{N}_a^v(\mathbf{y}))\prod\nolimits_{v'\in\mathcal{Y},v'\succ_a v}\left[1-\mathrm{Q}_a\left(v'|\mathrm{N}_a^{v'}(\mathbf{y})\right)\right].$$

Thus, the conditional choice probabilities of person a depend on the choices of others *only* via the number of people that select each option. This invariant restriction reduces the number of expressions we need to recover to

$$A\times Y\times\binom{A+Y-1}{Y}.$$

It follows that the number of moments in the data is larger than the number of expressions we want to recover if and only if

$$(Y+1)^A-1\geq A\times Y\times\binom{A+Y-1}{Y}.$$

When $Y=1$ and $A=2$, as in Example 1, then this inequality is not fulfilled. (Indeed, we explained earlier that the conditional choice probabilities are not identified from equilibrium behavior in this case.) However, in the case of one alternative (in addition to the default option) the inequality holds if there are at least five people in the network (that is, $A \geq 5$). Moreover, the number of people for which the inequality holds reduces to 4 whenever $Y \geq 2$. That is, a moderately large number of connections in the network seems to help the identification of the conditional choice probabilities. Of course, this analysis is still not complete as some of the moments in the data could still be linearly dependent.

4 Extension to Random Preferences

This section extends the previous model to allow for the possibility of randomness in preferences as well as in consideration sets. In this case, the choice rule $\mathrm{R}_a(\cdot|\,\mathcal{C})$ in Section 2 is not an indicator function but a distribution on $\mathcal{Y}$. We naturally let $\mathrm{R}_a(v|\,\mathcal{C}) = 0$ if $v \notin \mathcal{C}$.

Keeping unchanged the other parts of the model, the probability that person a selects (at the moment of choosing) alternative $v \in \mathcal{Y}$ is given by

$$\mathrm{P}_a(v|\mathbf{y}) = \sum\nolimits_{\mathcal{C}\subseteq 2^{\mathcal{Y}}} \mathrm{R}_a(v|\,\mathcal{C}) \prod\nolimits_{v'\in\mathcal{C}} \mathrm{Q}_a\left(v'|\mathrm{N}_a^{v'}(\mathbf{y})\right) \prod\nolimits_{v'\notin\mathcal{C}} \left(1 - \mathrm{Q}_a\left(v'|\mathrm{N}_a^{v'}(\mathbf{y})\right)\right). \tag{3}$$

The probability of selecting the default option o is (as before) $\prod_{v\in\mathcal{Y}} (1 - \mathrm{Q}_a(v|\mathrm{N}_a^v(\mathbf{y})))$.

Example 3: If we use the logit model to represent the random preferences of person a, then the probability that the person selects alternative 1 when alternative 2 is also part of her consideration set would be given by

$$\mathrm{R}_a(1|\,\{o,1,2\}) = \frac{\exp\left(\mathrm{U}_a^1\right)}{\exp\left(\mathrm{U}_a^1\right) + \exp\left(\mathrm{U}_a^2\right)}.$$

In this expression, U_a^1 and U_a^2 are the mean expected utilities that agent a gets from alternatives 1 and 2, respectively. ■

Under this alternative specification of the model, the identification of $(\mathrm{P}_a)_{a\in\mathcal{A}}$ follows from the same arguments. We will thereby assume the researcher knows the conditional choice probabilities, and focuses on recovering the set of connections, the profile of choice probabilities, and the attention mechanism. The main result is as follows.

Proposition 6 *Suppose A1-A3 are satisfied and we know* $(P_a)_{a\in\mathcal{A}}$. *Then, the set of connections* $\Gamma = (\mathcal{A}, e)$ *and the attention mechanism* $(Q_a)_{a\in\mathcal{A}}$ *are identified. For each* $a \in \mathcal{A}$*, the random preferences* R_a *are also identified if and only if, in addition, we have that* $|\mathcal{N}_a| \geq Y-1$.

Remark. The last result extends to the case in which the random preferences include the default option o with only one caveat. In this case the attention mechanism can be recovered up to ratios of the form $\mathrm{Q}_a\left(v|\mathrm{N}_a^v\left(\mathbf{y}\right)\right)/\mathrm{Q}_a\left(v|0\right)$. That is, we can only recover how much extra attention a person pays to each option as more of her friends select that option.

Here again, under assumption A1-A3, the choices of peers act as exclusion restrictions in the stochastic variation of the consideration sets and this variation suffices to recover the connections between the people in the network and the attention mechanism. The only difference from the case of deterministic preferences is that we need additional variation across consideration sets. The extra condition guarantees the matrix of coefficients for the R_a's in expression (3) is full rank. Moreover, we show that $|\mathcal{N}_a| \geq Y-1$ is not only sufficient, but necessary, to this end. We illustrate the last result by a simple example.

Example 2 (continued): Let us keep all the structure of Example 2 except for people's preferences, which we now assume are random. The identification of the set of connections and the attention mechanism follows from similar ideas. Thus we will only focus on recovering R_1, R_2, and R_3. Consider the next system of equations for Person 1

$$\begin{pmatrix} \mathrm{P}_1\left(1|y_1,o,o\right)/\mathrm{Q}_1\left(1|0\right) \\ \mathrm{P}_1\left(1|y_1,2,o\right)/\mathrm{Q}_1\left(1|0\right) \end{pmatrix} = \begin{pmatrix} 1-\mathrm{Q}_1\left(2|0\right) & \mathrm{Q}_1\left(2|0\right) \\ 1-\mathrm{Q}_1\left(2|1\right) & \mathrm{Q}_1\left(2|1\right) \end{pmatrix} \begin{pmatrix} \mathrm{R}_1\left(1|\left\{o,1\right\}\right) \\ \mathrm{R}_1\left(1|\left\{o,1,2\right\}\right) \end{pmatrix}$$

The fact that $\mathrm{R}_1\left(1|\left\{o,1\right\}\right)$ and $\mathrm{R}_1\left(1|\left\{o,1,2\right\}\right)$ can be recovered follows because, by A3, we have that

$$\det\begin{pmatrix} 1-\mathrm{Q}_1\left(2|0\right) & \mathrm{Q}_1\left(2|0\right) \\ 1-\mathrm{Q}_1\left(2|1\right) & \mathrm{Q}_1\left(2|1\right) \end{pmatrix} = \mathrm{Q}_1\left(2|1\right)-\mathrm{Q}_1\left(2|0\right) > 0.$$

The extra condition, $|\mathcal{N}_a| \geq Y - 1$, and A3 guarantee that the matrix of coefficients for the R_a's is always full column rank. ■

5 Illustration: Choosing a Restaurant

This section simulates a sequence of choices for a simple version of our initial model that we apply to a problem of choosing a restaurant. The exercise has two aims. First, we illustrate how people′s mistakes relate to the structure of the network. Second, we show how the main parts of the model can indeed be estimated from the sequence of choices that we simulate.

5.1 Simulation

There are five people in the network. Their reference groups are as follows

$$\mathcal{N}_1 = \{2\}, \mathcal{N}_2 = \{1\}, \mathcal{N}_3 = \{1, 2\}, \mathcal{N}_4 = \{5\}, \text{ and } \mathcal{N}_5 = \{4\}.$$

Note that each person has at least one friend, so A2 is satisfied. There are two possible restaurants at which individuals can have dinner. Restaurant 1 offers Mediterranean food and Restaurant 2 is a Steakhouse. Thus, $\mathcal{Y} = \{1, 2\}$. The default option o involves eating at home. The preferences of these people are as follows

$$2 \succ_1 1,\ 1 \succ_2 2,\ 2 \succ_3 1,\ 1 \succ_4 2, \text{ and } 1 \succ_5 2.$$

That is, Persons 2, 4, and 5 prefer Mediterranean food, and Persons 1 and 3 prefer the Steakhouse. We will assume the attention mechanism is invariant across people and alternatives. In this case, we can avoid some sub-indices and let $\mathrm{Q}(v|\mathrm{N}_a^v(\mathbf{y}))$ be the probability that person a pays attention to restaurant $v \in \mathcal{Y}$ if $\mathrm{N}_a^v(\mathbf{y})$ people of her reference group did so the last time they reviewed strategies. We initially let

$$\mathrm{Q}(v|0) = \frac{1}{4},\ \mathrm{Q}(v|1) = \frac{3}{4}, \text{ and } \mathrm{Q}(v|2) = \frac{7}{8}.$$

The rates for their Poisson "alarm clocks" are 1.

The equilibrium behavior of this restaurant choice model is a joint distribution μ with support on 243 choice configurations (3^5). We simulated a long sequences of choices and calculated the equilibrium behavior. (See the Appendix for more details.) From the equilibrium behavior we can easily obtain the marginal distributions across people.

$\mu_1(o) = 0.30$	$\mu_2(o) = 0.30$	$\mu_3(o) = 0.19$	$\mu_4(o) = 0.30$	$\mu_5(o) = 0.30$
$\mu_1(1) = 0.30$	$\mu_2(1) = 0.40$	$\mu_3(1) = 0.29$	$\mu_4(1) = 0.50$	$\mu_5(1) = 0.50$
$\mu_1(2) = 0.40$	$\mu_2(2) = 0.30$	$\mu_3(2) = 0.52$	$\mu_4(2) = 0.20$	$\mu_5(2) = 0.20$

Finally, from these marginals we get the following probabilities of making mistakes.

	Person 1	Person 2	Person 3	Person 4	Person 5
Probability of Mistakes	60%	60%	48%	50%	50%

Note that Persons 2 and 4 are identical in all respect except in the type of friend they have. In particular, Person 4 shares with her friend the same preferences; the opposite is true for Person 2. This difference leads Person 4 to make fewer mistakes. It becomes clear from this illustration that homophyly is good news in our model! In addition, note that Person 3, having more friends, makes also fewer mistakes.

To illustrate a bit more how the network structure shapes people's mistakes, let us add two more connections in the model. In particular, let us assume that Person 3 is part of the consideration sets of Persons 1 and 2. That is,

$$\mathcal{N}_1 = \{2, 3\}, \mathcal{N}_2 = \{1, 3\}, \mathcal{N}_3 = \{1, 2\}, \mathcal{N}_4 = \{5\}, \text{ and } \mathcal{N}_5 = \{4\}.$$

Repeating the previous exercise, the new network generates the following marginal distributions.

$\mu_1(o) = 0.12$	$\mu_2(o) = 0.12$	$\mu_3(o) = 0.12$	$\mu_4(o) = 0.30$	$\mu_5(o) = 0.30$
$\mu_1(1) = 0.21$	$\mu_2(1) = 0.42$	$\mu_3(1) = 0.22$	$\mu_4(1) = 0.50$	$\mu_5(1) = 0.50$
$\mu_1(2) = 0.67$	$\mu_2(2) = 0.46$	$\mu_3(2) = 0.66$	$\mu_4(2) = 0.20$	$\mu_5(2) = 0.20$

From these marginals, the probabilities of mistakes are as follows.

	Person 1	Person 2	Person 3	Person 4	Person 5
Probability of Mistakes	33%	54%	34%	50%	50%

Note that the probabilities of making mistakes decrease for Persons 1, 2, and 3. But the change is larger for Persons 1 and 3 as they share the same preferences over the restaurants.

5.2 Estimation

This section uses the sequence of choices we simulated in the previous section to show that the main parts of the model can indeed be estimated. (In this case, identification follows by Proposition 4.) To this end, we will use the second specification of the network structure. Also, to make the analysis more tractable, we will impose three extra conditions: First, we will assume each person has at most two friends. Second, we will assume the attention mechanism is invariant across people and alternatives. Third, we will let the network be undirected. Under these assumptions the number of possible sets of connections among people or networks is 112.[4] In addition, recall that there are 5 people in the population and two restaurants. Thus, the number of profiles of strict preferences $\succ = (\succ_a)_{a\in\mathcal{A}}$ is $2^5 = 32$. Finally, the attention mechanism has 3 parameters to estimate

$$\mathbf{Q} = \left(\mathrm{Q}\left(v|0\right), \mathrm{Q}\left(v|1\right), \mathrm{Q}\left(v|2\right)\right)'.$$

In line with A3, we will consider attention mechanisms that respect the monotonicity condition. That is, $\mathrm{Q}(v|0) < \mathrm{Q}(v|1) < \mathrm{Q}(v|2)$.[5] In addition, given A1, we let $\mathrm{Q}(v|\mathrm{N}_a^v(\mathbf{y})) \in (0,1)$ for $\mathrm{N}_a^v(\mathbf{y}) = 0, 1, 2$. Let us indicate by $\theta = (\Gamma, \succ, \mathbf{Q})$ an element in the space of possible parameters we want to estimate. Each of them induces a transition rate matrix $\mathcal{M}(\theta)$.

We normalize the intensity parameter λ_a to 1 for all $a \in \mathcal{A}$. Thus, for each θ, we can construct the transition rate matrix $\mathcal{M}(\theta)$ using equations (1) and (2). In turn, this information

[4] Without any restriction there are $2^{25} = 33,554,432$ possible network configurations.

[5] Technically speaking, for estimation purposes, we can only impose weak inequalities.

allows to calculate the so called transition matrix

$$\mathcal{P}(\theta, \Delta) = e^{\Delta \mathcal{M}(\theta)}.$$

We can use the latter to build the log-likelihood function

$$\mathrm{L}_T(\theta) = \Sigma_{t=0}^{T-1} \ln \mathcal{P}_{\iota(\mathbf{y}_t), \iota(\mathbf{y}_{t+1})}(\theta, \Delta)$$

where $\iota(\mathbf{y}) \in \left\{1, 2, ..., \overline{\mathcal{Y}}^A\right\}$ is the position of $\mathbf{y}$ according the lexicographic order, and $\mathcal{P}_{k,m}(\theta, \Delta)$ is the (k, m)-th element of the matrix $\mathcal{P}(\theta, \Delta)$. Finally, let us define the estimated parameters as follows

$$\widehat{\theta}_T = \arg\max_\theta \mathrm{L}_T(\theta).$$

For a sequence of $T = 15,000$ observations the Maximum Likelihood estimates are as follows

Network	$\widehat{\mathcal{N}}_1 = \{2\}, \widehat{\mathcal{N}}_2 = \{1\}, \widehat{\mathcal{N}}_3 = \{1,2\}, \widehat{\mathcal{N}}_4 = \{5\}$, and $\widehat{\mathcal{N}}_5 = \{4\}$
Preferences	$2 \widehat{\succ}_1 1$, $1 \widehat{\succ}_2 2$, $2 \widehat{\succ}_3 1$, $1 \widehat{\succ}_4 2$, and $1 \widehat{\succ}_5 2$
Attention Mechanism	$\widehat{\mathrm{Q}}(v\|0) = 0.26$, $\widehat{\mathrm{Q}}(v\|1) = 0.75$, and $\widehat{\mathrm{Q}}(v\|2) = 0.87$

.

In summary, the estimates correctly recover the set of connections and the strict preference order of each person in the network and closely approximates the attention mechanism. The Appendix contains a more extensive Monte Carlo study of the performance of our estimator. We estimated the attention mechanism assuming that the preference orders and the network structure are known. The estimator performs well in terms of the mean bias and the root mean squared error.

6 Final Remarks

This paper offers a new model of interdependent choices that combines the dynamic model of social interactions of Blume (1993, 1995) with the (single-agent) model of random consideration sets of Manzini and Mariotti (2014). From a theoretical perspective, we state equilibrium existence and characterize equilibrium behavior. We also illustrate how the network structure shapes people′s mistakes. From an applied perspective, in our model, the choices of peers act as exclusion restrictions in the stochastic variation of the considerations sets. This feature allows us to recover (from data) the main parts of the model without relying on variation of the set of alternative options or menus. Interestingly, we show that in addition of nonparametrically recovering the preference ranking of each person and the attention mechanism, we also identify the set of connections or edges between the people in the network.

7 Proofs

Proof of Proposition 1: For an irreducible, finite-state continuous Markov chain the steady-state μ exists and it is unique. Thus, we only need to prove that A1 implies that the Markov chain induced by our model is irreducible. First note that, under A1, for each $a \in \mathcal{A}$, $y_a \in \mathcal{Y}$, and $\mathbf{y} \in \overline{\mathcal{Y}}^A$, we have that

$$1 > \mathrm{P}_a \left(y_a | \mathbf{y} \right) = \mathrm{Q}_a \left(y_a | \mathrm{N}_a^{y_a} \left(\mathbf{y} \right) \right) \prod\nolimits_{v \in \mathcal{Y}, v \succ_a y_a} \left[1 - \mathrm{Q}_a \left(v | \mathrm{N}_a^v \left(\mathbf{y} \right) \right) \right] > 0.$$

To show irreducibility, let $\mathbf{y}$ and $\mathbf{y}'$ be two different choice configurations. It follows from expression (2) that we can go from one configuration to the other one in less than A steps with positive probability.

The characterization of μ follows as the invariant distribution satisfies the balance condition $\sum_{\mathbf{y}' \in \overline{\mathcal{Y}}^A} \mu \left(\mathbf{y}' \right) \mathrm{m}(\mathbf{y} \mid \mathbf{y}') = 0$ for each $\mathbf{y} \in \overline{\mathcal{Y}}^A$. The next steps show this claim.

$$\begin{aligned}
\sum\nolimits_{\mathbf{y}' \in \overline{\mathcal{Y}}^A} \mu \left(\mathbf{y}' \right) \mathrm{m} \left(\mathbf{y} \mid \mathbf{y}' \right) &= 0 \\
\mu \left(\mathbf{y} \right) \left(- \sum\nolimits_{\mathbf{y}' \in \overline{\mathcal{Y}}^A \setminus \{ \mathbf{y} \}} \mathrm{m} \left(\mathbf{y}' \mid \mathbf{y} \right) \right) + \sum\nolimits_{\mathbf{y}' \in \overline{\mathcal{Y}}^A \setminus \{ \mathbf{y} \}} \mu \left(\mathbf{y}' \right) \mathrm{m} \left(\mathbf{y} \mid \mathbf{y}' \right) &= 0 \\
-\mu \left(\mathbf{y} \right) \sum\nolimits_{a \in \mathcal{A}} \sum\nolimits_{y_a' \in \overline{\mathcal{Y}} \setminus \{ y_a \}} \lambda_a \mathrm{P}_a \left(y_a' | \mathbf{y} \right) + \sum\nolimits_{a \in \mathcal{A}} \sum\nolimits_{y_a' \in \overline{\mathcal{Y}} \setminus \{ y_a \}} \mu \left(y_a', \mathbf{y}_{-a} \right) \lambda_a \mathrm{P}_a \left(y_a | y_a', \mathbf{y}_{-a} \right) &= 0 \\
-\mu \left(\mathbf{y} \right) \sum\nolimits_{a \in \mathcal{A}} \lambda_a \left(1 - \mathrm{P}_a \left(y_a | \mathbf{y} \right) \right) + \sum\nolimits_{a \in \mathcal{A}} \sum\nolimits_{y_a' \in \overline{\mathcal{Y}} \setminus \{ y_a \}} \mu \left(y_a', \mathbf{y}_{-a} \right) \lambda_a \mathrm{P}_a \left(y_a | y_a', \mathbf{y}_{-a} \right) &= 0 \\
\frac{1}{\sum_{a \in \mathcal{A}} \lambda_a} \sum\nolimits_{a \in \mathcal{A}} \lambda_a \left\{ \sum\nolimits_{y_a' \in \overline{\mathcal{Y}}} \mu \left(y_a', \mathbf{y}_{-a} \right) \mathrm{P}_a \left(y_a | y_a', \mathbf{y}_{-a} \right) \right\} &= \mu \left(\mathbf{y} \right) \\
\frac{1}{\sum_{a \in \mathcal{A}} \lambda_a} \sum\nolimits_{a \in \mathcal{A}} \lambda_a \mathrm{P}_a \left(y_a | \mathbf{y} \right) \mu_{-a} \left(\mathbf{y}_{-a} \right) &= \mu \left(\mathbf{y} \right).
\end{aligned}$$

In moving from the fifth line to the sixth one we used the fact that, in our model, $\mathrm{P}_a \left(y_a | y_a', \mathbf{y}_{-a} \right) = \mathrm{P}_a \left(y_a | \mathbf{y}_{-a} \right)$ for any $y_a' \in \overline{\mathcal{Y}}^A$. ■

Proof of Proposition 2: By A1, P_a has full support for all $\mathbf{y}$. By A2 and A3, $\mathrm{P}_a \left(v | \mathbf{y} \right)$ is strictly decreasing in $\mathrm{N}_a^{v'} \left(\mathbf{y} \right)$ for each $v' \succ_a v$. Thus, we can recover $\mathcal{N}_a$. Since this is true for each $a \in \mathcal{A}$, we can get $\Gamma = \left(\mathcal{A}, e \right)$. In addition, from variation in $\mathrm{N}_a^{v'} \left(\mathbf{y} \right)$ for each $v' \neq v$, we can recover person a's upper level set that corresponds to option v. That is,

$$\left\{ v' \in \mathcal{Y} : v' \succ_a v \right\}.$$

By repeating the exercise with each alternative, we can recover $\succ_a$. Finally, suppose that y_a^* is the most preferred alternative for person a. Then,

$$\mathrm{P}_a\left(y_a^*|\mathbf{y}\right)=\mathrm{Q}_a\left(y_a^*|\mathrm{N}_a^{y_a^*}\left(\mathbf{y}\right)\right).$$

It follows that we can recover $\mathrm{Q}_a\left(y_a^*|\mathrm{N}_a^{y_a^*}\left(\mathbf{y}\right)\right)$. By proceeding in descending preference ordering we can then recover $\mathrm{Q}_a\left(v|\mathrm{N}_a^{v}\left(\mathbf{y}\right)\right)$ for all $v\in\mathcal{Y}$. ■

Proof of Proposition 3: Since $\lim_{\Delta\to 0}\mathcal{P}\left(\Delta\right)=\mathcal{M}$, we can recover transition rate matrix from the data. Recall that

$$\mathrm{m}\left(\mathbf{y}'\mid\mathbf{y}\right)=\begin{cases}0 & \text{if } \sum_{a\in\mathcal{A}}1\left(y_a'\neq y_a\right)>1\\ \sum_{a\in\mathcal{A}}\lambda_a\mathrm{P}_a\left(y_a'|\mathbf{y}\right)1\left(y_a'\neq y_a\right) & \text{if } \sum_{a\in\mathcal{A}}1\left(y_a'\neq y_a\right)=1\end{cases}.$$

Thus, $\lambda_a\mathrm{P}_a\left(y_a'|\mathbf{y}\right)=\mathrm{m}(y_a',\mathbf{y}_{-a}\mid\mathbf{y})$. It follows that we can recover $\lambda_a\mathrm{P}_a\left(v|\mathbf{y}\right)$ for each $v\in\overline{\mathcal{Y}}$, $\mathbf{y}\in\overline{\mathcal{Y}}^A$, and $a\in\mathcal{A}$. Note that, for each $\mathbf{y}\in\overline{\mathcal{Y}}^A$,

$$\sum\nolimits_{v\in\overline{\mathcal{Y}}}\lambda_a\mathrm{P}_a\left(v|\mathbf{y}\right)=\lambda_a\sum\nolimits_{v\in\overline{\mathcal{Y}}}\mathrm{P}_a\left(v|\mathbf{y}\right)=\lambda_a.$$

Then we can also recover λ_a for each $a\in\mathcal{A}$. ■

Proof of Proposition 4: This proof builds on Theorem 1 of Blevins (2017) and Theorem 3 of Blevins (2018). For the present case, it follows from the last two theorems, that the transition rate matrix $\mathcal{M}$ is generically identified if, in addition to the conditions in Proposition 4, we have that

$$(Y+1)^A-AY-1\geq\frac{1}{2}.$$

This condition is always satisfied if $A>1$. Identification of $\mathcal{M}$ follows because, by A2, $A\geq 2$. We can then uniquely recover $\left(\mathrm{P}_a\right)_{a\in\mathcal{A}}$ from $\mathcal{M}$. See the proof of Proposition 3 ■

Proof of Proposition 5: From Proposition 1, μ satisfies

$$\mu\left(\mathbf{y}\right)=\frac{1}{A}\sum\nolimits_{a\in\mathcal{A}}\mathrm{P}_a\left(y_a|\mathbf{y}\right)\mu_{-a}\left(\mathbf{y}_{-a}\right)\ \text{ for each }\mathbf{y}\in\overline{\mathcal{Y}}^A. \tag{4}$$

We only need to show that if $(\mathrm{P}_a)_{a\in A}$ is a set of compatible conditional distributions, then μ = P solves (4). If we let μ = P, then right hand side of (4) is

$$\frac{1}{A}\sum\nolimits_{a\in\mathcal{A}}\mathrm{P}_a\left(y_a|\mathbf{y}\right)\sum\nolimits_{v\in\overline{\mathcal{Y}}}\mathrm{P}\left(v,\mathbf{y}_{-a}\right)=\frac{1}{A}\sum\nolimits_{a\in\mathcal{A}}\mathrm{P}\left(\mathbf{y}\right)=\frac{A}{A}\mathrm{P}\left(\mathbf{y}\right)=\mathrm{P}\left(\mathbf{y}\right).$$

In addition, the left hand side of (4) is

$$\mu\left(\mathbf{y}\right)=\mathrm{P}\left(\mathbf{y}\right).$$

Thus $\mu\left(\mathbf{y}\right)=\mathrm{P}(\mathbf{y})$ solves (4) for each $\mathbf{y}\in\overline{\mathcal{Y}}^{A}$. ■

Proof of Proposition 6: Note that expression (3) can be rewritten as follows

$$\mathrm{P}_a\left(v|\mathbf{y}\right)=\sum\nolimits_{\mathcal{C}\subseteq\mathcal{Y}}\mathrm{R}_a\left(v|\ \mathcal{C}\right)\prod\nolimits_{v'\in\mathcal{C}}\mathrm{Q}_a\left(v'|\mathrm{N}_a^{v'}\left(\mathbf{y}\right)\right)\prod\nolimits_{v''\notin\mathcal{C}}\left(1-\mathrm{Q}_a\left(v''|\mathrm{N}_a^{v''}\left(\mathbf{y}\right)\right)\right)=$$
$$\mathrm{Q}_a\left(v|\mathrm{N}_a^{v}\left(\mathbf{y}\right)\right)\sum\nolimits_{\mathcal{C}\subseteq\mathcal{Y},v\in\mathcal{C}}\mathrm{R}_a\left(v|\ \mathcal{C}\right)\prod\nolimits_{v'\in\mathcal{C}\setminus\{v\}}\mathrm{Q}_a\left(v'|\mathrm{N}_a^{v'}\left(\mathbf{y}\right)\right)\prod\nolimits_{v''\notin\mathcal{C}}\left(1-\mathrm{Q}_a\left(v''|\mathrm{N}_a^{v''}\left(\mathbf{y}\right)\right)\right).$$

Thus, by A2 and A3, we can state whether $a'\in\mathcal{N}_a$ by checking whether $\mathrm{P}_a\left(v|y_1=o,...,y_A=o\right)$ moves up when we change $y_{a'}$ from o to v for some v in $\mathcal{A}$. It follows that the network structure is identified.

Let $\mathbf{y}$ be such that $\mathrm{N}_a^{v}\left(\mathbf{y}\right)=0$ and let us assume that at least one person (different from a) in $\mathbf{y}$ selected the default option (i.e., there is at least one $y_{a'}=o$ with $a'\neq a$). Let $\mathbf{y}'$ be such that

$$\mathrm{N}_a^{v'}\left(\mathbf{y}\right)=\mathrm{N}_a^{v'}\left(\mathbf{y}'\right)\ \text{ for all } v'\neq v \text{ and } \mathrm{N}_a^{v}\left(\mathbf{y}\right)=1.$$

Note that

$$\mathrm{P}_a\left(v|\mathbf{y}'\right)/\mathrm{P}_a\left(v|\mathbf{y}\right)=\mathrm{Q}_a\left(v|1\right)/\mathrm{Q}_a\left(v|0\right).$$

Also

$$\mathrm{P}_a\left(o|\mathbf{y}'\right)/\mathrm{P}_a\left(o|\mathbf{y}\right)=\left(1-\mathrm{Q}_a\left(v|1\right)\right)/\left(1-\mathrm{Q}_a\left(v|0\right)\right).$$

Thus, by A3, $\mathrm{Q}_a\left(v|0\right)$ and $\mathrm{Q}_a\left(v|1\right)$ can be recovered from the data. By implementing a similar procedure for different values of $\mathrm{N}_a^{v}\left(\mathbf{y}\right)$ we can recover Q_a. Finally, since this is true for any arbitrary a, then we can recover the attention mechanism $\left(\mathrm{Q}_a\right)_{a\in\mathcal{A}}$.

We finally show that R_a is identified if and only if (in addition to A1-A3) we have that $|\mathcal{N}_a| \geq Y - 1$. We will present the idea for $v = 1$ and agent a. (The proof immediately extends to other agents and alternatives.) We want to recover $\mathrm{R}_a(1|\,\mathcal{C})$ for all $\mathcal{C}$. To simplify the exposition we will write

$$
\begin{aligned}
|\mathcal{N}_a| &= N \\
\mathrm{Q}_a(v|m) &= \mathrm{Q}^1(v|m) \\
1 - \mathrm{Q}_a(v|m) &= \mathrm{Q}^0(v|m)
\end{aligned}
$$

We have a set of equations indexed by $\mathbf{y}$

$$
\mathrm{P}_a(1|\mathbf{y}) / \mathrm{Q}_a\left(1|\mathrm{N}_a^1(\mathbf{y})\right) = \sum\nolimits_{\mathcal{C}\subseteq\mathcal{Y}, v\in\mathcal{C}} \mathrm{R}_a(v|\,\mathcal{C}) \prod\nolimits_{k\in\mathcal{C}\setminus\{v\}} \mathrm{Q}^1\left(k|\mathrm{N}_a^k(\mathbf{y})\right) \prod\nolimits_{k\notin\mathcal{C}\setminus\{v\}} \mathrm{Q}^0\left(k|\mathrm{N}_a^k(\mathbf{y})\right).
$$

To present the ideas more clear let $A(N,Y)$ be the matrix of coefficients in front of the $\mathrm{R}_a^{\prime}$s. The above system of equations has a unique solution if and only if $A(N,Y)$ has full column rank. The column of $A(N,Y)$ that corresponds to any given $\mathcal{C} \subseteq \mathcal{Y}$ consists of the elements of the following form

$$
\prod\nolimits_{k\in\mathcal{Y}} \mathrm{Q}^{1(k\in\mathcal{C})}\left(k|\mathrm{N}^k\right)
$$

where $\mathrm{N}^k \in \{0,1,...,N\}$ and $\sum_k \mathrm{N}^k \leq N$. The last claim in the proposition follows from the next lemma.

Lemma 1: Assume that A1-A3 hold. For all $Y \geq 2$ and $N \geq 1$

$$
N \geq Y - 1 \Longleftrightarrow A(N,Y) \text{ has full column rank.}
$$

Proof.

Step 1. To illustrate how the idea works, note that $A(1,2)$ and $A(1,3)$ can be written as follows

$$
A(1,2) = \begin{pmatrix} \mathrm{Q}^0(2|0) & \mathrm{Q}^1(2|0) \\ \mathrm{Q}^0(2|1) & \mathrm{Q}^1(2|1) \end{pmatrix}
$$

and

$$A(1,3) = \begin{pmatrix} Q^0(3|0)\,Q^0(2|0) & Q^0(3|0)\,Q^1(2|0) & Q^1(3|0)\,Q^0(2|0) & Q^1(3|0)\,Q^1(2|0) \\ Q^0(3|0)\,Q^0(2|1) & Q^0(3|0)\,Q^1(2|1) & Q^1(3|0)\,Q^0(2|1) & Q^1(3|0)\,Q^1(2|1) \\ Q^0(3|1)\,Q^0(2|0) & Q^0(3|1)\,Q^1(2|0) & Q^1(3|1)\,Q^0(2|0) & Q^1(3|1)\,Q^1(2|0) \end{pmatrix}$$

$$= \begin{pmatrix} Q^0(3|0)\,A(1,2) & Q^1(3|0)\,A(1,2) \\ Q^0(3|1)\,A(0,2) & Q^1(3|1)\,A(0,2) \end{pmatrix}$$

where $A(0,2) = \begin{pmatrix} Q^0(2|0) & Q^1(2|0) \end{pmatrix}$.

Similarly, the matrix $A(N, Y+1)$ can be written as follows

$$A(N,Y+1) = \begin{pmatrix} Q^0(Y+1|0)\,A(N,Y) & Q^1(Y+1|0)\,A(N,Y) \\ Q^0(Y+1|1)\,A(N-1,Y) & Q^1(Y+1|1)\,A(N-1,Y) \\ Q^0(Y+1|2)\,A(N-2,Y) & Q^1(Y+1|2)\,A(N-2,Y) \\ \dots & \dots \\ Q^0(Y+1|N)\,A(0,Y) & Q^1(Y+1|N)\,A(0,Y) \end{pmatrix}.$$

Note that $A(K,Y)$ is a sub-matrix of $A(K+1,Y)$ for all K (with the same number of columns). Thus, it is clear that $A(N,Y+1)$ has full column rank only if $A(N,Y)$ and $A(N-1,Y)$ have both full column rank, which is the same as to say $A(N-1,Y)$ has full column rank. We next show that under A3, if $A(N-1,Y)$ has full column rank, then $A(N,Y+1)$ has full column rank too. To this end, let M be a matrix obtained deleting rows from $A(N-1,Y)$ in such a way that $\det(M) > 0$. Then, by A3, we have

$$\det \begin{pmatrix} Q^0(Y+1|0)\,M & Q^1(Y+1|0)\,M \\ Q^0(Y+1|1)\,M & Q^1(Y+1|1)\,M \end{pmatrix} = \left(Q^1(Y+1|1) - Q^1(Y+1|0)\right)^{2^{Y-1}} \det(M)^2 > 0.$$

In summary, we have that

$$A(N,Y+1) \text{ has full column rank } \Longleftrightarrow A(N-1,Y) \text{ has full column rank.}$$

Step 2. Consider $(N,Y) = (1,2)$. Note that

$$A(1,2) = \begin{pmatrix} Q^0(2|0) & Q^1(2|0) \\ Q^0(2|1) & Q^1(2|1) \end{pmatrix}$$

has full column rank since $\det(A(1,2)) = \mathrm{Q}^1(2|1) - \mathrm{Q}^1(2|0) > 0$. In addition, any $A(N,2)$ with $N \geq 1$ will have full column rank because $A(1,2)$ is a sub-matrix of $A(N,2)$ with the same number of columns.

Finally, note that $A(1,3)$ has not full column rank since the number of columns is higher than the number of rows.

Step 3. From Step 1 we got that, for all $Y \geq 2$ and $N \geq 1$, we have that

$$A(N, Y+1) \text{ has full column rank } \Longleftrightarrow A(N-1, Y) \text{ has full column rank.}$$

From Step 2, we get that $A(N,2)$ (with $N \geq 1$) has full column rank and $A(1,3)$ has not full column rank. The claim in Lemma 1 follows by combining these three results. ■

References

[1] Abaluck, J., and A. Adams (2017): "What do consumers consider before they choose? identification from asymmetric demand responses," Mimeo.

[2] Aguiar, V. (2017): "Random categorization and bounded rationality," *Economics Letters*, 159, 46–52.

[3] Aguiar, V., M. Boccardi, M. Dean (2016): "Satisficing and stochastic choice," *Journal of Economic Theory*, 166, 445–482.

[4] Aguiar, V., M. Boccardi, N. Kashaev, and J. Kim (2019): "Does random consideration explain behavior when choice is hard? evidence from a large-scale experiment," Mimeo.

[5] Allen, B. (1982): "Some Stochastic Processes of Interdependent Demand and Technological Diffusion of an Innovation Exhibiting Externalities Among Adopters," *International Economic Review*, 23, 595–608.

[6] Arnold, B., and S. Press (1989): "Compatible conditional distributions," *Journal of the American Statistical Association*, 84, 152–156.

[7] Barseghyan, L., F. Molinari, and M. Thirkettle (2019): "The nature of risk preferences: Evidence from insurance choices," Mimeo.

[8] Barry, A., and S. Press (1989): "Compatible Conditional Distributions," *Journal of the American Statistical Association*, 84, 152–156.

[9] Blevins, J. (2018): "Identification and Estimation of Continuous Time Dynamic Discrete Choice Games," Mimeo.

[10] Blume, L. (1993): "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 5, 387–424.

[11] Blume, L. (1995): "The Statistical Mechanics of Best-Response Strategy Revision," *Games and Economic Behavior*, 11, 111–45.

[12] Blume, L., W. Brock, S. Durlauf, and R. Jayaraman (2015): "Linear social interactions models," *Journal of Political Economy*, 123, 444–496.

[13] Blume, L., and S. Durlauf (2003): "Equilibrium Concepts for Social Interaction Models," *International Game Theory Review*, 5, 193–209.

[14] Bonaldi, P., A. Hortaçsu, and J. Kastl (2015): "An empirical analysis of funding costs spillovers in the euro-zone with application to systemic risk," Technical report, National Bureau of Economic Research.

[15] Borah, A., and C. Kops (2018): "Choice via Social Influence," Mimeo.

[16] Brady, R., and J. Rehbeck (2016): "Menu-dependent stochastic feasibility," *Econometrica*, 84, 1203–1223.

[17] Caplin, A., M. Deanz, and J. Leahy (2018): "Rational Inattention, Optimal Consideration Sets and Stochastic Choice," Mimeo.

[18] Caplin, A., J. Leahy, and F. Matějka (2016): "Rational Inattention and Inference from Market Share Data," Mimeo.

[19] Cattaneo, M., Y. Masatlioglu, and E. Suleymanov (2017): "A random attention model," Mimeo.

[20] Chambers C. P., C. Cuhadaroglu, and Y. Masatlioglu (2019): "Behavioral influence," Mimeo.

[21] Conlon, C., and J. Mortimer (2013): "Demand estimation under incomplete product availability,".*American Economic Journal: Microeconomics*, 5, 1–30.

[22] De Paula, Á., I. Rasul, and P. Souza (2018): "Recovering social networks from panel data: identification, simulations and an application," Mimeo.

[23] Draganska, M., and D. Klapper (2011): "Choice set heterogeneity and the role of advertising: Analysis with micro and macro data," *Journal of Marketing Research*, 48, 653–669.

[24] Gaynor, M., C. Propper, and S. Seiler (2016): "Free to choose? reform, choice, and consideration sets in the English national health service," *American Economic Review*, 106, 3521–3557.

[25] Goeree, M. (2008): "Limited information and advertising in the us personal computer industry," *Econometrica*, 76, 1017–1074.

[26] Horan, S. (2018): "Random consideration and choice: a case study of default options," Technical report.

[27] Kaiser, M., and N. Cressie (2000): "The Construction of Multivariate Distributions from Markov Random Fields," *Journal of Multivariate Analysis*, 73, 199–220.

[28] Lazzati, N. (2018): "Co-Diffusion of Technologies in Social Networks," Mimeo.

[29] Lleras, J., Y. Masatlioglu, D. Nakajima, and E. Y. Ozbay (2017): "When more is less: Limited consideration," *Journal of Economic Theory*, 170, 70–85.

[30] Manresa, E. (2013): "Estimating the structure of social interactions using panel data," Mimeo.

[31] Manski, C. (1977): "The structure of random utility models," *Theory and Decision*, 8, 229–54.

[32] Manzini, P., and M. Mariotti (2014): "Stochastic Choice and Consideration Sets," *Econometrica*, 82, 1153–1176.

[33] McFadden, D. (1973): "Conditional Logit Analysis of Qualitative Choice Behavior," Frontiers in Econometrics, Academic Press: New York, 105–142.

[34] Mehta, N., S. Rajiv, and K. Srinivasan (2003): "Price uncertainty and consumer search: A structural model of consideration set formation," *Marketing Science*, 22, 58–84.

[35] Roberts, J., and J. Lattin (1991): "Development and testing of a model of consideration set composition," *Journal of Marketing Research*, 28, 429–440.

[36] Rook, D. (2016): "Popular Influence on Stochastic Choice," Mimeo.

[37] Wichmann, B., M. Chen, and W. Adamowicz (2016): "Social Networks and Choice Set Formation in Discrete Choice Models," *Econometrics*, 4, 1–26.

Appendix: Simulation for Section 5

This appendix describes how we generated the observations for the restaurant model

Let $\lambda = \sum_{a \in \mathcal{A}} \lambda_a$. We generate the data according to an iterative procedure for a fixed time period $\mathcal{T}$. The k-th iteration of the procedure is as follows:

(i) Given $\mathbf{y}_{k-1}$ set $\mathbf{y}_k = \mathbf{y}_{k-1}$;

(ii) Generate a draw from the exponential distribution with mean $1/\lambda$ and call it x_k;

(iii) Randomly sample an agent from the set $\mathcal{A}$, such that the probability that a is picked is λ_a/λ;

(iv) Given the agent selected in the previous step and the current choice configuration $\mathbf{y}_k$ construct a consideration set using Q_a;

(v) If the consideration set is empty, then set $y_{a,k} = 0$. Otherwise pick the best alternative according to the preference order of agent a from the consideration set and assign it to $y_{a,k}$.

Given the initial configuration of choices $\mathbf{y}_0$ we applied the above algorithm till we reached $\sum_k x_k > \mathcal{T}$ (On average the length of the sequence is $\lambda\mathcal{T}$). Define $z_k = \sum_{l \leq k} x_l$. The continuous time data is $\{(y_k, z_k)\}$. The discrete time data is obtained from the continuous time data by splitting the interval $[0, \mathcal{T}]$ into $T = [\mathcal{T}/\Delta]$ intervals and recording the configuration of the network at every time period $t = i\Delta$, $i = 0, 1, ..., [\mathcal{T}/\Delta]$.

To elaborate a bit more on the estimation of the attention mechanism, we also generated 1000 data samples over the period of $[0, 25000]$ and then for $\Delta \in \{5/3, 5/2, 5, 25\}$ constructed discrete data sets (the sample size $\{1000, 5000, 10000, 15000\}$). The network structure and the preference orders were assumed to be known. (Its values are as the ones in Section 5.) So only optimization over the consideration probabilities was performed. The results of the simulations are presented in the next table.

Table 1. Bias and Root Mean Squared Error (RMSE) ($\times 10^{-3}$)

Sample Size		$\mathrm{Q}\,(v\|0)$	$\mathrm{Q}\,(v\|1)$	$\mathrm{Q}\,(v\|2)$
1000	Bias	42.3	18.0	15.1
	RMSE	43.3	20.9	17.8
5000	Bias	9.4	2.7	0.3
	RMSE	10.2	5.6	4.4
10000	Bias	4.7	-0.6	-0.5
	RMSE	5.5	3.5	3.3
15000	Bias	3.5	-1.3	2.8
	RMSE	4.2	3.3	-0.3

Notes: The sample sizes correspond to $\Delta = 25$ for the sample size 1000, $\Delta = 5$ for the sample size 5000, $\Delta = 5/2$ for the sample size 10000, and $\Delta = 5/3$ for the sample size 15000. The number of replications is 1000.