

# Peer Effects in Random Consideration Sets\*

Nail Kashaev	Natalia Lazzati
Western University	UC Santa Cruz
nkashaev@uwo.ca	nlazzati@ucsc.edu

This version: May 17, 2021

First version: April 13, 2019, arXiv:1904.06742

**Abstract** We develop a dynamic model of discrete choice that incorporates peer effects into random consideration sets. We characterize the equilibrium behavior and study the empirical content of the model. In our setup, changes in the choices of friends affect the distribution of the consideration sets. We exploit this variation to recover the ranking of preferences, attention mechanisms, and network connections. These nonparametric identification results allow unrestricted heterogeneity across people and do not rely on the variation of either covariates or the set of available options. Our methodology leads to a maximum-likelihood estimator that performs well in simulations. We apply our results to an experimental dataset that has been designed to study the visual focus of attention.

JEL codes: C31, C33, D83, O33

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

---

\*For their useful comments and suggestions we would like to thank Victor Aguiar, Roy Allen, Tim Conley, Paola Manzini, Marco Mariotti, and Salvador Navarro. We would also like to thank the audience at Workshop in Experiments, Revealed Preferences, and Decisions, the Inter University Colloquium on Analytical Research in Economics, the CARE-Colloquium series at University of Kansas, NAWM ES 2021, and the seminar presentation at El Colegio de Mexico.

# 1. Introduction

In the last few years, the basic rational choice model of decision-making has been revised in different ways, partly recognizing that cognitive factors might play an important role in determining the choices of people. These revisions have produced, among others, the consideration set models. In these models, people do not consider all the available options at the moment of choosing, but a subset of them. It is still an open question in the literature how the consideration sets are formed and what their main determinants are. We think that peer effects might be of first-order importance in the formation process of the consideration sets.<sup>1</sup> To study this possibility, this paper builds a dynamic model where the choices of friends affect the subset of options that a person ends up considering. We show the empirical content of the model and apply our main results to an experimental dataset that has been designed to evaluate the visual focus of attention of people. In our application, peer effects have a large impact on the attention mechanism.

In the dynamic model we build, people are linked through a social network. At a randomly given time, a person gets the opportunity to select a new option out of a finite set of alternatives. The person sticks to her new option until the revision opportunity arises again. People do not consider all the available options at the moment of revising their selection. Instead, each person first forms a consideration set and then picks her most preferred option from it. The distinctive feature of the model is that the probability that a given alternative enters the consideration set depends on the number of friends that are currently adopting that option. This theoretical model leads to a sequence of choices that evolves through time according to a Markov random process. The dynamic system induced by our model has a unique equilibrium that specifies the fraction of time that each vector of joint choices prevails in the long run.

In the consideration set models, the relative frequency by which a person chooses each alternative does not necessarily respect her ranking of preference. Formally, let us say that a person makes a mistake when, at the moment of choosing, she selects an alternative that is dominated by another option according to her preference order. In our model, a person could seldom choose her most preferred option at equilibrium. The set of connections, or the network structure, shapes the nature and the strength of the mistakes that people make. Thus, our model can be thought of as a model of peer effects in mistakes. We show via example that having friends with similar preferences—a phenomenon known as homophily—reduces the frequency by which people make mistakes. In addition, the example shows that having more friends also results in fewer mistakes.

Having established equilibrium existence and characterized equilibrium behavior, we consider a researcher who observes a long sequence of choices made by the members of the network. We show that all primitives of the model can be uniquely recovered. These primitives include the ranking of preferences, the attention mechanism (or consideration probabilities), and the network structure. There are two aspects of our nonparametric identification results that deserve special attention.

---

<sup>1</sup>This possibility has been (explicitly or implicitly) discussed by other researchers in specific contexts—e.g., the choices of peers may help us discover a new television show (Godes and Mayzlin, 2004), a new welfare program (Caeyers et al., 2014), a new retirement plan (Duflo and Saez, 2003), a new restaurant (Qiu et al., 2018), or an opportunity to protest (Enikolopov et al., 2020).

First, we allow unrestricted heterogeneity across people with respect to all parts of the model. Second, in contrast to most other works on consideration sets, we do not rely on the variation of either covariates or the set of available options (or menu) to recover these components.

In our dynamic model, the observed choices of network members are generated by a system of conditional choice probabilities: each of these conditional choice probabilities specifies the distribution of choices of a given person conditional on the choices of others (at the moment of revising her selection). The identification strategy we offer is a two-step procedure. First, we show how to identify the primitives of the model using these conditional choice probabilities. Second, we study identification of the conditional choice probabilities from observed data.

To recover the primitives of the model from the conditional choice probabilities we exploit that, in our framework, changes in the choices of friends induce stochastic variation in the consideration set probabilities. We use this variation to recover the set of connections between the people in the network and their ranking of preferences. We 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. As we explained earlier, this identification result is nonparametric and allows for full heterogeneity across people regarding preferences and attention probabilities.

To identify the conditional choice probabilities, we consider two datasets: continuous-time data and discrete-time data with arbitrary time intervals. These two datasets coincide in that they provide a long sequence of choices from people in the network. They differ in the timing at which the researcher observes these choices. In continuous-time datasets, the researcher observes people’s choices in real time. We can think of this dataset as the “ideal dataset.” With the proliferation of online platforms and scanners, this sort of data might be available in some applications. In this dataset, the researcher directly recovers the conditional choice probabilities. In discrete-time datasets, the researcher observes the joint configuration of choices at fixed time intervals (e.g., the choice configuration is observed every Monday). In this case, the conditional choice probabilities are not directly observed or recovered, they need to be inferred from the data. Adding an extra mild condition, we show that the conditional choices are also uniquely identified. For this last result we invoke insights from [Blevins \(2017, 2018\)](#). Briefly, the reason for which the second dataset suffices to recover the primitives of the model lies on the fact that the transition rate matrix of a continuous-time process with independent revision times across people is rather parsimonious. In particular, the probability that two or more people revise their selected options at the same time is zero. As a consequence, the transition rate matrix has zeros in many known locations. The non-zero elements can then be nicely recovered and they constitute a one-to-one mapping with the conditional choice probabilities.

Our initial results rely on a few simplifying restrictions. In particular, the preferences of each person are deterministic; we assume that there exists one alternative (the default) that is picked if and only if nothing else is considered; and we let the distribution of consideration sets be multiplicatively separable across alternatives and the probability of including each option depends on the number (but not the identity) of the friends that selected that option. We extend our model in various directions to relax these assumptions. First, we consider random preferences (on top of

random consideration sets). The network structure and the attention mechanism are identified without any new assumptions. The random preferences are identified if and only if each person has “enough” friends —the actual number of friends needed depends on the number of options. Second, we analyze a model with no default. All the primitives of this model are identified if there are more than three options in the set of available alternatives. Finally, we consider a set of extensions where consideration sets are formed arbitrarily (e.g., no multiplicative separability assumption). In these extensions, different friends may have different effects on the attention mechanism of a given person. In this model, we still can uniquely recover preferences and the network structure. The consideration probabilities, in general, are partially identified. However, we show that different forms of symmetry between consideration set probabilities are sufficient for their identification.

After presenting the main identification results, we propose a consistent maximum-likelihood estimator of all parameters of the model that uses discrete-time data. Our Monte Carlo simulations imply that our estimator performs well even for relatively small datasets.


Last, but not least, we apply our model to the experimental dataset used by Bai et al. (2019), which has been designed to evaluate the visual focus of attention of people. In the experiment, a group of people seated around a circle were asked to play the Resistance game. This is a party game in which people communicate with each other to find the deceptive players. A tablet was placed in front of each player and it recorded the direction of the sight of the player every third of a second (i.e., we consider each player looking to her left, right, or center). The structural model we offer allows us to disentangle the effects of two key forces in the visual focus of attention of people: directional sight preferences and peer effects in gazing. Regarding directional sight preferences, our estimates provide very strong support to the left-to-right bias. That is, the observation that people tend to look to the left more often than to the right.<sup>2</sup> (This observation has guided, among others, the design of many online platforms!) Regarding peer effects in gazing, our results support the observation that people redirect their visual attention by following the gaze orientation of others, a phenomenon called “gaze following.”<sup>3</sup>

We finally relate our results with the existing literature. From a modeling perspective, our setup combines the dynamic model of social interactions of Blume (1993, 1995) with the (single-agent) model of random consideration sets of Manski (1977) and Manzini and Mariotti (2014). By adding peer effects in the consideration sets we can use variation in the choices of others as the main tool to recover preferences. 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. (2019), Cattaneo et al. (2020), Horan (2019), Lleras et al. (2017), Manzini and Mariotti (2014), and Masatlioglu et al. (2012). (See Aguiar et al., 2021 for a comparison of several consideration set models in an experimental setting.) Other papers have relied on the existence of exogenous covariates that shift preferences or consideration sets. The latter include Barseghyan et al. (2021b,a), Crawford et al. (2021), Conlon

---

<sup>2</sup>See, for example, the experimental findings of Spalek and Hammad (2004, 2005) and Reutskaja et al. (2011). The left-to-right bias has also been documented in marketing research <https://www.nngroup.com/articles/horizontal-attention-original-research/>. See also Maass and Russo (2003) and reference therein.

<sup>3</sup>See, for example, Gallup et al. (2012).

and Mortimer (2013), Draganska 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 (2021) via an approach that exploits symmetry breaks with respect to the full consideration set model. Aguiar  Kashaev (2021), Crawford et al. (2021), and Dardanoni et al. (2020) use repeated choices but do not allow for peer effects (i.e., they work with panel data).

There is a vast econometric literature on identification of models of social interactions where choices of peers affect preferences but not the choice sets (see Blume et al., 2011, Bramoullé et al., 2020, De Paula, 2017, and Graham, 2015 for comprehensive reviews of this literature). We depart from this literature in that in our framework the direct interdependence between choices (endogenous effects) is captured by consideration sets (choice sets), not preferences.<sup>4</sup> As we mentioned earlier, 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 identifies the network structure but in their model peers do not affect consideration sets but directly change preferences (among other differences).

In the paper, we connect the equilibrium behavior of our model with the Gibbs equilibrium. This connection is similar to the one in Blume and Durlauf (2003). We also distinguish our approach with the more standard models of peer effects in preferences. To this end we embed the models of Brock and Durlauf (2001, 2002) into our dynamic revision process.

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 (2020) 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, the main assumptions, and some key insights of our approach (including how it differs from the more standard approach of peer effects in preferences). Section 3 describes the equilibrium behavior. Section 4 studies the empirical content of the model. Section 5 extends the initial idea to contemplate random preferences (besides random consideration sets), the case of no-default option, and more general formation process for the consideration sets. Section 7 applies our model to experimental data on visual focus of attention. Section 8 concludes, and all the proofs are collected in Appendix A. Appendices B and C cover the Gibbs random field model and provide some simulation results, respectively. Appendix D contains additional results for our empirical application.

---

<sup>4</sup>Implicitly, preferences in our model capture the so-called contextual effects – effects of predetermined factors that are not affected by peers.

## 2. The Model

This section describes the model and the main assumptions we invoke in the paper. It also compares our approach to the more standard model of peer effects in preferences.

### 2.1. Social Network, Consideration Sets, and Choices

**Network and Choice Configuration** 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  $\mathcal{A} = \{1, 2, \dots, A\}$  is the finite 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:

$$\mathcal{N}_a = \{a' \in \mathcal{A} : a' \neq a \text{ and there is an edge from } a \text{ to } a' \text{ in } e\}.$$

There is a set of alternatives  $\bar{\mathcal{Y}} = \mathcal{Y} \cup \{o\}$ , where  $\mathcal{Y} = \{1, 2, \dots, Y\}$  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.<sup>5</sup> We extend the analysis to random preferences in Section 5.1 and fully relax the specification of the default option in Section 5.2. We refer to a vector  $\mathbf{y} = (y_a)_{a \in \mathcal{A}} \in \bar{\mathcal{Y}}^A$  as a choice configuration.

**Choice Revision** We model the revision of choices as a standard continuous-time Markov process. In particular, we assume that people are endowed with independent Poisson alarm clocks with rates  $\lambda = (\lambda_a)_{a \in \mathcal{A}}$ .<sup>6</sup> At randomly given moments (exponentially distributed with mean  $1/\lambda_a$ ) the alarm of Person  $a$  goes off.<sup>7</sup> When this happens, the person selects the most preferred alternative among the ones she is actually considering. Formally, if  $\mathcal{C} \subseteq \mathcal{Y}$  is her consideration set, then the choice of Person  $a$  can be represented by an indicator function

$$R_a(v | \mathcal{C}) = \mathbb{1}(v \succ_a v' \text{ for all } v' \in \mathcal{C} \text{ and } v \in \mathcal{C})$$

that takes value 1 if  $v$  is the most preferred option in  $\mathcal{C}$  according to  $\succ_a$ . If, at the moment of choosing, the consideration set of Person  $a$  does not include any alternative in  $\mathcal{Y}$ , then the person selects the default option.

**Peer Effects in the Formation of Consideration Sets** In our model, whether Person  $a$  pays attention to a particular alternative depends on her own choice and the configuration of choices

<sup>5</sup>Although we assume the same default option for all people, all our results go through if people have different default options as far as the researcher knows the identity of the worst option for each person.

<sup>6</sup>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.

<sup>7</sup>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  $1/\lambda_a$ . These differences are independent across people and time.

of her friends at the moment of revising her selection. We indicate by  $Q_a(v | \mathbf{y})$  the probability that Person  $a$  pays attention to alternative  $v$  given a choice configuration  $\mathbf{y}$ . It follows that the probability of facing consideration set  $\mathcal{C}$  has the form of

$$\prod_{v \in \mathcal{C}} Q_a(v | \mathbf{y}) \prod_{v \notin \mathcal{C}} (1 - Q_a(v | \mathbf{y})).$$

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

$$P_a(v | \mathbf{y}) = Q_a(v | \mathbf{y}) \prod_{v' \in \mathcal{Y}, v' \succ_a v} (1 - Q_a(v' | \mathbf{y})). \quad (1)$$

The default option is assumed to be always considered. Since it is also assumed to be the worst for every person, the default option is chosen by the person if, and only if, nothing else is considered. Thus, the probability of selecting  $o$  is just  $\prod_{v \in \mathcal{Y}} (1 - Q_a(v | \mathbf{y}))$ . Leaving aside peer effects, this process of formation of consideration sets is analogous to the one studied by [Manski \(1977\)](#) and [Manzini and Mariotti \(2014\)](#). In [Section 5.3](#) we extend our analysis to a more general setting.

Altogether, the three elements just described characterize our initial model of peer effects in random consideration sets. (As mentioned above, we consider later several extensions.) This model leads to a sequence of joint choices that evolves through time according to a Markov random process. An equilibrium in this model is an invariant distribution on the space of joint choices of the dynamic process, that describes the frequency by which each member of the group selects each option. We next present three key assumptions we invoke in the paper.

## 2.2. Main Assumptions

Our results build on three assumptions. Let  $N_a^v(\mathbf{y})$  be the number of friends of Person  $a$  who select option  $v$  in choice configuration  $\mathbf{y}$ . Formally,

$$N_a^v(\mathbf{y}) = \sum_{a' \in \mathcal{N}_a} \mathbb{1}(y_{a'} = v).$$

We indicate by  $|\mathcal{N}_a|$  the cardinality of  $\mathcal{N}_a$ . The three assumptions are as follows.

**(A1)** For each  $a \in \mathcal{A}$ ,  $v \in \mathcal{Y}$ , and  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ ,  $1 > Q_a(v | \mathbf{y}) > 0$ .

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

**(A3)** For each  $a \in \mathcal{A}$ ,  $v \in \mathcal{Y}$ , and  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ ,

$$Q_a(v | \mathbf{y}) \equiv Q_a(v | y_a, N_a^v(\mathbf{y})) \text{ is strictly increasing in } N_a^v(\mathbf{y}).$$

Assumption A1 states that, for any choice configuration, the probability of considering each option is strictly positive and lower than one, independently on how many friends have selected

that option. This assumption captures the idea that people can eventually pay attention to an alternative for various reasons that are outside the control of our model (e.g., watching an ad on television or receiving a coupon). It also allows people to eventually disregard any further consideration of a given option, including the one that she is currently adopting. Assumption A1 guarantees that each subset of options is (ex-ante) considered with nonzero probability. Assumption A2 requires each person to have at least one friend. Assumption A3 states that the probability that a given person pays attention to a specific option depends on the current choice of the person and the number (but not the identity) of friends that currently selected it. Assumption A3 also states that each person pays more attention to a particular option if more of her friends are adopting it.

### 2.3. Insights of the Model

In our model, the probability that Person  $a$  selects option  $v$  (at the moment of choosing) given network configuration  $\mathbf{y}$  takes the form of

$$P_a(v | \mathbf{y}) = Q_a(v | y_a, N_a^v(\mathbf{y})) \prod_{v' \in \mathcal{Y}, v' \succ_a v} (1 - Q_a(v' | y_a, N_a^{v'}(\mathbf{y}))).$$

Let us formally say a person makes a mistake whenever she selects a dominated option. Note that the preference order of Person  $a$  over the available alternatives is fixed. The choices of friends simply redirect the attention of the person from one alternative to another one. As a consequence, at any given time, the probability that Person  $a$  selects a low ranked alternative that is popular among her friends can be much higher than the probability of selecting her most preferred option. In summary, in our model, the network structure does not shape preferences (or payoffs across alternatives) but the strength of the mistakes that people make. We show later, via example, that having similar friends – a phenomenon known as homophily – can sharply reduce the frequency of mistakes that the person makes. Having more friends acts in the same direction.

Our model markedly differs from the more standard model of peer effects in preferences. To illustrate some of the main differences, note that in the same dynamic revision process we can easily embed a model of peer effects in preferences by modifying the conditional choice probabilities. Let us assume that individual choices are path independent, in the sense that choosing from a large set of options can be broken up into choosing from smaller subsets (Luce’s Choice Axiom). If we add peer effects in preferences into the resulting choice model, then the probability that option  $v$  is chosen given network configuration  $\mathbf{y}$  can be expressed as follows

$$P_a(v | \mathbf{y}) = U_a^v(y_a, N_a^v(\mathbf{y})) / (\sum_{v' \in \bar{\mathcal{Y}}} U_a^{v'}(y_a, N_a^{v'}(\mathbf{y}))),$$

where  $U_a^v(y_a, N_a^v(\mathbf{y}))$  is the utility Person  $a$  gets from alternative  $v$  given her previous choice  $y_a$  and the number of friends who also picked it,  $N_a^v(\mathbf{y})$ .<sup>8</sup> In this model, the choices of friends affect

---

<sup>8</sup>If we further assume, as in Brock and Durlauf (2001, 2002), that the payoff Person  $a$  gets from alternative  $v$  is the sum of a deterministic private utility the person receives from the choice,  $\delta_{a,v}$ , a social component that depends on the number of friends that select that option,  $S_{a,v}(\mathbf{y}) \equiv S_{a,v}(y_a, N_a^v(\mathbf{y}))$ , and an idiosyncratic term that



the value that Person  $a$  assigns to each option. Thus, at each given time, the relative frequency by which Person  $a$  selects option  $v$  over option  $v'$  reflects her relative expected utility of the first option with respect to the second one. It follows that the alternative with the highest expected utility is always selected more often. This is in sharp contrast with what happens in our peer effects model in consideration sets, as we explained above.

Let us finally remark that, expressed as we did, these two models also have different empirical implications. In particular, in the random consideration set model, the probability that Person  $a$  selects alternative  $v$  does not vary if some of her friends switch from the default option to any option that is dominated by alternative  $v$  according to  $\succ_a$ . In contrast, in the model of peer effects in preferences, the probability of selecting alternative  $v$  strictly increases under the same scenario. It follows that if we can recover the conditional choice probabilities from the data (as we do in Section 4), then the two models can be set apart.

### 3. Equilibrium Behavior

The independent and identically distributed Poisson alarm clocks, which lead the selection revision process, guarantee that at each moment of time at most one person revises her selection almost surely. Thus, the transition rates between choice configurations that differ in more than one person changing the current selection are zero. The advantage of this fact for model identification is that there are fewer terms to recover. Blevins (2017, 2018) offers a nice discuss of this feature and its advantage over discrete time models. Formally, the transition rate from choice configuration  $\mathbf{y}$  to any different one  $\mathbf{y}'$  is as follows

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

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 diagonal terms are simply build from these other values as follows

$$m(\mathbf{y} | \mathbf{y}) = - \sum_{\mathbf{y}' \in \bar{\mathcal{Y}}^A \setminus \{\mathbf{y}\}} m(\mathbf{y}' | \mathbf{y}).$$

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

---

is doubly exponentially distributed with index parameter 1, then  $U_a^v(y_a, N_a^v(\mathbf{y}))$  takes the familiar exponential form  $\exp(\delta_{a,v} + S_{a,v}(y_a, N_a^v(\mathbf{y})))$ .

$\mathcal{M}_{11} = m \left( (o, o, \dots, o)' \mid (o, o, \dots, o)' \right)$ . Formally, let  $\iota(\mathbf{y}) \in \{1, 2, \dots, (Y+1)^A\}$  be the position of  $\mathbf{y}$  according to the lexicographic order. Then,

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

An equilibrium in this model is an invariant distribution  $\mu : \bar{\mathcal{Y}}^A \rightarrow [0, 1]$ , with  $\sum_{\mathbf{y} \in \bar{\mathcal{Y}}^A} \mu(\mathbf{y}) = 1$ , 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

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

The next proposition establishes equilibrium existence and uniqueness for our model.

**Proposition 3.1.** *If Assumption A1 is satisfied, then there exists a unique equilibrium  $\mu$ .*

**Remark.** In Appendix B we relate the equilibrium notion we use with the so-called Gibbs equilibrium. This alternative solution concept has been extensively used in Economics to study social interactions. (See Allen, 1982, Blume, 1993, 1995, and Blume and Durlauf, 2003, among many others.) The existence of Gibbs equilibrium requires extra symmetry restrictions among people. While these additional restrictions make the model more tractable, they are hard to justify in empirical applications. This is why we selected a more general approach.

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

**Example 1.** The network consists of two identical people that select between two alternatives, option 1 and the default option  $o$ . The rates for their Poisson alarm clocks are 1. Let us also assume, to simplify the setup, that the probability of paying attention to a particular option only depends on the current choice of the other person (i.e.,  $Q_a(v \mid y_a, N_a^v(\mathbf{y})) = Q_a(v \mid N_a^v(\mathbf{y}))$ ) and is the same for both people (i.e.,  $Q(\cdot) = Q_1(\cdot) = Q_2(\cdot)$ ). Thus, for  $a = 1, 2$ , we get that

$$P_a(1 \mid \mathbf{y}) = Q(1 \mid N_a^v(\mathbf{y})) \text{ and } P_a(o \mid \mathbf{y}) = 1 - Q(1 \mid N_a^v(\mathbf{y})).$$

The transition rate matrix  $\mathcal{M}$  is as follows.

	$(o, o)$	$(o, 1)$	$(1, o)$	$(1, 1)$
$(o, o)$	$-2Q(1 \mid 0)$	$Q(1 \mid 0)$	$Q(1 \mid 0)$	0
$(o, 1)$	$1 - Q(1 \mid 0)$	$-1 + Q(1 \mid 0) - Q(1 \mid 1)$	0	$Q(1 \mid 1)$
$(1, o)$	$1 - Q(1 \mid 0)$	0	$-1 + Q(1 \mid 0) - Q(1 \mid 1)$	$Q(1 \mid 1)$
$(1, 1)$	0	$1 - Q(1 \mid 1)$	$1 - Q(1 \mid 1)$	$-2 + 2Q(1 \mid 1)$

The transition rate matrix  $\mathcal{M}$  is naturally more complex when there are more alternatives or more people. However, the structure of the zeros in  $\mathcal{M}$  is rather similar. In particular, there are many zeros in known locations. As we mentioned earlier, this feature of the model is particularly attractive for identification and estimation.

The invariant distribution of choices, or steady-state equilibrium, satisfies  $\mu\mathcal{M} = \mathbf{0}$ . Solving this system of equations, we get that the steady-state equilibrium is

$$\begin{aligned}\mu(o, o) &= \frac{[1 - Q(1 | 0)][1 - Q(1 | 1)]}{1 - Q(1 | 1) + Q(1 | 0)}, \\ \mu(o, 1) = \mu(1, o) &= \frac{Q(1 | 0)[1 - Q(1 | 1)]}{1 - Q(1 | 1) + Q(1 | 0)}, \\ \mu(1, 1) &= \frac{Q(1 | 0)Q(1 | 1)}{1 - Q(1 | 1) + Q(1 | 0)}.\end{aligned}$$

The steady-state equilibrium is a joint distribution on the pair of choice configurations. It states the fraction of time that each pair of choices  $(y_1, y_2)$  prevails in the long run. ■

The second example shows how the strength of the mistakes that people make relates to the structure of the network. We will use the simulated data in the example later to show how the main parts of the model can, indeed, be estimated from a sequence of choices.

**Example 2.** There are five people in the network. Their reference groups are as follows

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

Note that each person has at least one friend, so Assumption A2 is satisfied. Moreover, the network is directed since  $3 \notin \mathcal{N}_1, \mathcal{N}_2$  and  $1, 2 \in \mathcal{N}_3$ . There are two possible alternatives  $\mathcal{Y} = \{1, 2\}$  in addition to the default option  $o$ . The preferences of these people are as follows

$$2 \succ_1 1, \quad 1 \succ_2 2, \quad 2 \succ_3 1, \quad 1 \succ_4 2, \quad 1 \succ_5 2.$$

That is, Persons 2, 4, and 5 prefer option 1, and Persons 1 and 3 prefer option 2. We will assume the attention mechanism is invariant across people and alternatives. To keep the exercise simple, as in Example 1, we will also assume the probability of paying attention to a given option only depends on the choices of friends. In this case, we indicate by  $Q(v | N_a^v(\mathbf{y}))$  the probability that Person  $a$  pays attention to option  $v \in \mathcal{Y}$  if the number of friends that are currently selecting that option is  $N_a^v(\mathbf{y})$ . We initially let

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

The rates for their Poisson alarm clocks are 1.

The equilibrium behavior of this choice model is a joint distribution  $\mu$  with support on 243 choice configurations ( $3^5$ ). We simulated a long sequence of choices and calculated the equilibrium behavior. (See Appendix C for more details.) From the equilibrium distribution we can obtain the marginal equilibrium probabilities for each person. Each of these marginals specifies the fraction of time that a given person selects each alternative in the long run. These marginal probabilities are shown in Table 1. It is interesting to observe that, for some people in the network, the relative

**Table 1** – Marginal Equilibrium Probabilities for Directed Network

$\mu_1(o) = 0.29$	$\mu_2(o) = 0.29$	$\mu_3(o) = 0.19$	$\mu_4(o) = 0.30$	$\mu_5(o) = 0.30$
$\mu_1(1) = 0.30$	$\mu_2(1) = 0.40$	$\mu_3(1) = 0.30$	$\mu_4(1) = 0.50$	$\mu_5(1) = 0.50$
$\mu_1(2) = 0.41$	$\mu_2(2) = 0.31$	$\mu_3(2) = 0.51$	$\mu_4(2) = 0.20$	$\mu_5(2) = 0.20$

Notes: These probabilities are estimated using simulated data, sample size = 15000.

**Table 2** – Probabilities of Mistakes for Directed Network

	Person 1	Person 2	Person 3	Person 4	Person 5
Probability of Mistakes	59%	60%	49%	50%	50%

Notes: These probabilities are computed using the estimated marginal equilibrium probabilities from Table 1.

frequency of choosing each alternative does not respect their order of preferences. In particular, notice that Persons 4 and 5 select the default more often than option 2, which dominates the former according to their preferences.

As before, let us say that a person makes a mistake when does not select her most preferred alternative. From the equilibrium marginals, we can calculate the probabilities of making mistakes. These probabilities are displayed in Table 2. Note that Persons 2 and 4 are identical in all respects 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. Hence, having friends with similar preferences (homophily) seems to help each person to consider more often her best alternative. Also, note that Person 3 has more friends and makes fewer mistakes. Thus, having more friends is also helpful in this application.

To show a bit more how the network structure shapes people’s mistakes, we add two more connections in the model. In particular, let us assume that Person 3 is a friend of Persons 1 and 2. That is,

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

In this case, the network becomes undirected. Repeating the previous exercise, the new network generates the marginal equilibrium distributions depicted in Table 3. From these marginals, the probabilities of mistakes are presented on Table 4. The probabilities of making mistakes decrease for Persons 1, 2, and 3. But notice that the change is larger for Persons 1 and 3 as they share the same preferences over the alternatives. ■

**Table 3** – Marginal Equilibrium Probabilities for Undirected Network

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

Notes: These probabilities are estimated using simulated data, sample size = 15000.

**Table 4** – Probabilities of Mistakes for Undirected Network

	Person 1	Person 2	Person 3	Person 4	Person 5
Probability of Mistakes	33%	58%	34%	50%	50%

Notes: These probabilities are computed using the estimated marginal equilibrium probabilities from Table 3.

## 4. Empirical Content of the Model

This section provides conditions under which the researcher can uniquely recover (from a long sequence of choices) the set of connections  $\Gamma = (\mathcal{A}, e)$ , the profile of strict preferences  $\succ = (\succ_a)_{a \in \mathcal{A}}$ , the attention mechanism  $Q = (Q_a)_{a \in \mathcal{A}}$ , and the rates of the Poisson alarm clocks  $\lambda = (\lambda_a)_{a \in \mathcal{A}}$ .

We separate the identification analysis in two parts. Let  $P = (P_a)_{a \in \mathcal{A}}$  be the profile of choice probabilities of people in the network. Each  $P_a(v | \mathbf{y}) : \bar{\mathcal{Y}} \times \bar{\mathcal{Y}}^{\mathcal{A}} \rightarrow (0, 1)$  specifies the (ex-ante) probability that Person  $a$  selects option  $v$  when the choice configuration is  $\mathbf{y}$ . Recall that, in our model,

$$P_a(v | \mathbf{y}) = Q_a(v | \mathbf{y}) \prod_{v' \in \mathcal{Y}, v' \succ_a v} (1 - Q_a(v' | \mathbf{y})),$$

and the probability of selecting the default option  $o$  is  $\prod_{v \in \mathcal{Y}} (1 - Q_a(v | \mathbf{y}))$ . First, we show that each set of conditional choice probabilities  $P$  maps into a different set of connections, profile of strict preferences, and attention mechanism. Thus, knowledge of the set of the conditional choice probabilities allows us to uniquely recover all the elements of the model. Second, we build identification of the conditional choice probabilities  $P$  from a long sequence of choices.

### 4.1. Identification of the Model from $P$

Under Assumptions A1 and A2, changes in the choices of friends induce stochastic variation of the consideration sets. Assumption A3 guarantees this variation is monotonic in the sense that the probability of considering one option increases with the number of friends that are currently adopting it. This stochastic variation in choices allows us to recover the set of connections between the people in the network and the ranking of preferences of each of them. We then sequentially identify the attention mechanism of each person moving from the most preferred alternative to the least preferred one. Proposition 4.1 presents our first identification result.

**Proposition 4.1.** *Under Assumptions A1-A3, the set of connections  $\Gamma$ , the profile of strict preferences  $\succ$ , and the attention mechanism  $Q$  are point identified from  $P$ .*

The next example sheds some light on the identification strategy in Proposition 4.1.

**Example 3.** Suppose there are three people  $\mathcal{A} = \{1, 2, 3\}$  that select between 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. Let  $\mathbf{y}$  be such that  $y_1 = o$ . The probability that Person 1 selects the default option  $o$  (given a profile of choices  $\mathbf{y}$  with  $y_1 = o$ ) is

$$P_1(o | \mathbf{y}) = \left(1 - Q_1\left(1 | o, N_1^1(\mathbf{y})\right)\right) \left(1 - Q_1\left(2 | o, N_1^2(\mathbf{y})\right)\right).$$

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

$$P_1(o | o, o, o) > P_1(o | o, 1, o).$$

Indeed, if  $2 \in \mathcal{N}_1$ , then the probability of choosing the default option by Person 1 should decrease if Person 2 picks something else. Also, if  $2 \notin \mathcal{N}_1$ , then the probability of choosing the default option by Person 1 should be invariant to choices of Person 2. Similarly,  $3 \in \mathcal{N}_1$  if and only if  $P_1(o | o, o, o) > P_1(o | o, o, 1)$ . Thus, we can learn the set of friends of Person 1 from observed  $P_1$ . Let us assume that  $\mathcal{N}_1 = \{2\}$ . To recover the preferences of Person 1 note that

$$\begin{aligned} P_1(1 | \mathbf{y}) &= Q_1\left(1 | o, N_1^1(\mathbf{y})\right) && \text{if } 1 \succ_1 2 \\ P_1(1 | \mathbf{y}) &= Q_1\left(1 | o, N_1^1(\mathbf{y})\right) \left(1 - Q_1\left(2 | o, N_1^2(\mathbf{y})\right)\right) && \text{if } 2 \succ_1 1 \end{aligned}$$

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

$$P_1(1 | o, o, o) > P_1(1 | o, 2, o).$$

Suppose that, indeed, we get that  $2 \succ_1 1$ . We can finally recover the attention mechanism (for  $y_1 = o$ ) via the next four probabilities in the data

$$\begin{aligned} P_1(2 | o, o, o) &= Q_1(2 | o, 0) && P_1(2 | o, 2, o) = Q_1(2 | o, 1) \\ P_1(1 | o, o, o) &= Q_1(1 | o, 0) (1 - Q_1(2 | o, 0)) && P_1(1 | o, 1, o) = Q_1(1 | o, 1) (1 - Q_1(2 | o, 0)) \end{aligned}$$

By considering two other choice profiles  $\mathbf{y}$  with  $y_1 = 1$  and  $y_1 = 2$  (instead of  $y_1 = o$ ), respectively, we can fully recover the attention mechanism of Person 1. By a similar exercise we can recover the sets of friends, preferences, and the attention mechanisms for Persons 2 and 3.  $\blacksquare$

## 4.2. Identification of P

This section studies identification of the conditional choice probabilities,  $P$ , and the rates of the Poisson alarm clocks from two different datasets. These two datasets coincide in that they contain long sequences of choices from people in the network. They differ in the timing at which the researcher observes these choices. In Dataset 1 people's choices are observed in real time. This allows the researcher to record the precise moment at which a person revises her strategy and the configuration of choices at that time. In Dataset 2 the researcher simply observes the joint configuration of choices at fixed time intervals.

Let us assume the researcher observes people's choices at time intervals of length  $\Delta$  and can consistently estimate  $\Pr(\mathbf{y}^{t+\Delta} = \mathbf{y}' | \mathbf{y}^t = \mathbf{y})$  for each pair  $\mathbf{y}', \mathbf{y} \in \bar{\mathcal{Y}}^A$ . 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 the transition rate matrix  $\mathcal{M}$  described in Equation (2) is given by

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

where  $e^{(\Delta\mathcal{M})}$  is the matrix exponential of  $\Delta\mathcal{M}$ . The two datasets we consider differ regarding  $\Delta$ : in Dataset 1 we let the time interval be very small. This is an ideal dataset that registers people’s choices at the exact time in which any given person revises her choice. As we mentioned earlier, with the proliferation of online platforms and scanner this sort of data might indeed be available for some applications. In Dataset 2 we allow the time interval to be of arbitrary size. The next table formally describes Datasets 1 and 2

**Dataset 1** The researcher knows  $\lim_{\Delta \rightarrow 0} \mathcal{P}(\Delta)$

**Dataset 2** The researcher knows  $\mathcal{P}(\Delta)$

In both cases, the identification question is whether (or under what extra restrictions) it is possible to uniquely recover  $\mathcal{M}$  from  $\mathcal{P}(\Delta)$ . The first result in this section is as follows.

**Proposition 4.2** (Dataset 1). *The conditional choice probabilities  $\mathbf{P}$  and the rates of the Poisson alarm clocks  $\lambda$  are identified from Dataset 1.*

The proof of Proposition 4.2 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 assuming  $\Delta \rightarrow 0$ . One of them requires the length of the interval  $\Delta$  to be below a threshold  $\bar{\Delta}$ . The main difficulty of this identification approach is that the value of the threshold depends on the details of the model that are unknown to the researcher. The second case requires the researcher to observe the dynamic system at two different intervals  $\Delta_1$  and  $\Delta_2$  that are not multiples 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 4.3** (Dataset 2). *If Assumption A2 is satisfied, the researcher knows  $\lambda$ , and  $\mathcal{M}$  has distinct eigenvalues that do not differ by an integer multiple of  $2\pi i/\Delta$ , where  $i$  here denotes the imaginary unit, then the conditional choice probabilities  $\mathbf{P}$  are generically identified from Dataset 2.*

The key element in proving Proposition 4.3 is that the transition rate matrix in 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 (see Example 1). It is important to remark that this aspect of the model is not shared by models of discrete-time revision processes, where people in the network revise choices simultaneously at fixed time intervals.

## 5. Extensions of the Model

### 5.1. Random Preferences

This section extends the initial model to allow randomness in preferences and in consideration sets. In this case, the choice rule  $R_a(\cdot | \mathcal{C})$  from Section 2 is not an indicator function but a distribution on  $\mathcal{Y}$ . We naturally let  $R_a(v | \mathcal{C}) = 0$  if  $v \notin \mathcal{C}$ . Note that the relation between preferences and consideration sets is completely unrestricted.

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

$$P_a(v | \mathbf{y}) = \sum_{\mathcal{C} \subseteq \mathcal{Y}} R_a(v | \mathcal{C}) \prod_{v' \in \mathcal{C}} Q_a(v' | y_a, N_a^{v'}(\mathbf{y})) \prod_{v' \notin \mathcal{C}} (1 - Q_a(v' | y_a, N_a^{v'}(\mathbf{y}))). \quad (3)$$

The probability of selecting the default option  $o$  is (as before)  $\prod_{v \in \mathcal{Y}} (1 - Q_a(v | y_a, N_a^v(\mathbf{y})))$ . The next example illustrates the random choice rule with the well-known logit model.

**Example 4.** 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

$$R_a(1 | \{1, 2\}) = \frac{\exp(U_a^1)}{\exp(U_a^1) + \exp(U_a^2)}.$$

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

Under this variant of the initial model, the identification of  $P$  follows directly from Propositions 4.2 and 4.3. We will thereby focus on recovering the set of connections, the choice rule, and the attention mechanism from  $P$ . The main result is as follows.

**Proposition 5.1.** *Suppose Assumptions A1-A3 are satisfied. Then, the set of connections  $\Gamma$  and the attention mechanism  $Q$  are point identified from  $P$ . For each  $a \in \mathcal{A}$ , the random preferences  $R_a$  are also point 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  $Q_a(v | y_a, N_a^v(\mathbf{y})) / Q_a(v | y_a, 0)$ . That is, we can only recover how much *extra* attention a person pays to each option as more of her friends select that option.

As in our previous results, under Assumptions A1-A3, observed variation in the choices of friends induce 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 with respect to the case of deterministic preferences is that with random preferences we need a larger number of friends for each person, i.e.,  $|\mathcal{N}_a| \geq Y - 1$  for each  $a \in \mathcal{A}$ . The extra condition



guarantees the matrix of coefficients for the  $R'_a$ s in expression (3) is full column rank. Indeed, 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 extension of Example 3 above.

**Example 3 (continued).** Let us keep all the structure of Example 3 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 following system of equations for Person 1 that connects the observed  $P$  with  $Q$  and  $R$

$$\begin{pmatrix} P_1(1 | y_1, o, o) / Q_1(1 | y_1, 0) \\ P_1(1 | y_1, 2, o) / Q_1(1 | y_1, 0) \end{pmatrix} = \begin{pmatrix} 1 - Q_1(2 | y_1, 0) & Q_1(2 | y_1, 0) \\ 1 - Q_1(2 | y_1, 1) & Q_1(2 | y_1, 1) \end{pmatrix} \begin{pmatrix} R_1(1 | \{o, 1\}) \\ R_1(1 | \{o, 1, 2\}) \end{pmatrix}.$$

The fact that  $R_1(1 | \{1\})$  and  $R_1(1 | \{1, 2\})$  can be recovered follows because, by Assumption A3, we have that the determinant of the matrix of the coefficients in the above system of equations is different from zero:

$$\det \begin{pmatrix} 1 - Q_1(2 | y_1, 0) & Q_1(2 | y_1, 0) \\ 1 - Q_1(2 | y_1, 1) & Q_1(2 | y_1, 1) \end{pmatrix} = Q_1(2 | y_1, 1) - Q_1(2 | y_1, 0) > 0.$$

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

## 5.2. No Default Option

In the initial model, the default option plays a special role: it is chosen if, and only if, nothing else is considered. In some settings, such default option may not exist.<sup>9</sup> This section offers a variant of the model that accommodates to this possibility.<sup>10</sup>

Let us assume there is no default option  $o$ , so that  $\bar{\mathcal{Y}} = \mathcal{Y}$ . The formation process of the consideration set is as before except that, since there is no default option, we need to specify what people do when the consideration set is empty. Given the dynamic nature of our model, we will simply assume that each person sticks to her previous choice if no alternative receives further consideration. Formally, the probability that Person  $a$  selects (at the moment of choosing) alternative  $v \in \mathcal{Y}$  is given by

$$P_a(v | \mathbf{y}) = Q_a(v | \mathbf{y}) \prod_{v' \in \mathcal{Y}, v' \succ_a v} (1 - Q_a(v' | \mathbf{y})) + 1(v = y_a) \prod_{v' \in \mathcal{Y}} (1 - Q_a(v' | \mathbf{y})). \quad (4)$$

**Proposition 5.2.** *Suppose that Assumptions A1-A3 are satisfied and  $Y \geq 3$ . Then, the set of connections  $\Gamma$ , the profile of strict preferences  $\succ$ , and the attention mechanism  $Q$  are point identified from  $P$ .*

<sup>9</sup>Also, see Horan (2019).

<sup>10</sup>For alternative ways to close the model see, for example, Barseghyan et al. (2021b).

### 5.3. More General Peer Effects in Consideration Sets

In our initial model, the probability that Person  $a$  faces consideration set  $\mathcal{C}$  given a choice configuration  $\mathbf{y}$  takes the form of

$$\prod_{v \in \mathcal{C}} Q_a(v | y_a, N_a^v(\mathbf{y})) \prod_{v \notin \mathcal{C}} (1 - Q_a(v | y_a, N_a^v(\mathbf{y}))).$$

This approach entails a multiplicative separable specification of alternatives in the consideration sets. It also assumes that the probability of considering an option depends only on the total number of friends who pick that option, but not on the identity of these friends (i.e., the effect of friends' choices is symmetric across friends). We next show that neither of these assumptions is essential for our approach. Indeed, all previous insights can be used here to extend the initial results.

For each Person  $a$  and configuration  $\mathbf{y}$ , let  $\eta_a(\cdot | \mathbf{y})$  be an attention index function from  $2^{\mathcal{Y}}$  to the positive reals. The value  $\eta_a(\mathcal{C} | \mathbf{y})$  captures the attention that Person  $a$  pays to the set of alternatives  $\mathcal{C} \in 2^{\mathcal{Y}}$  given the choice configuration  $\mathbf{y}$ . The attention-index measures how enticing a consideration set is (see [Aguiar et al. \(2021\)](#) for further details). We define the probability of facing consideration set  $\mathcal{C}$  as

$$\frac{\eta_a(\mathcal{C} | \mathbf{y})}{\sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} | \mathbf{y})}.$$

In this model, the consideration set probabilities are as in [Brady and Rehbeck \(2016\)](#). Since  $\eta_a$  can be identified only up to scale, we normalize the attention-index for the empty set to be 1 (i.e.,  $\eta(\emptyset | \mathbf{y}) = 1$  for all  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ ).<sup>11</sup> This more general setting covers our initial model, which is based on [Manzini and Mariotti \(2014\)](#), as a special case.

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

$$P_a(v | \mathbf{y}) = \sum_{\mathcal{C} \in 2^{\mathcal{Y}}: v \in \mathcal{C}} R_a(v | \mathcal{C}) \frac{\eta_a(\mathcal{C} | \mathbf{y})}{\sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} | \mathbf{y})}.$$

The probability of selecting the default option  $o$  is just  $\eta_a(\emptyset | \mathbf{y}) / \sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} | \mathbf{y})$ . To better understand the connection between our initial model and this extension note that, for any  $v \in \mathcal{Y}$  and  $\mathcal{C} \subseteq \mathcal{Y}$  such that  $v \notin \mathcal{C}$ , we have that

$$\frac{\eta_a(\mathcal{C} \cup \{v\} | \mathbf{y})}{\eta_a(\mathcal{C} | \mathbf{y})} = \frac{Q_a(v | \mathbf{y})}{1 - Q_a(v | \mathbf{y})}.$$

Thus,

$$Q_a(v | \mathbf{y}) = \frac{\eta_a(\mathcal{C} \cup \{v\} | \mathbf{y})}{\eta_a(\mathcal{C} \cup \{v\} | \mathbf{y}) + \eta_a(\mathcal{C} | \mathbf{y})}.$$

Based on this alternative specification, we accommodate Assumptions A1 and A3 as follows.

**(A1')** For each  $a \in \mathcal{A}$ ,  $v \in \mathcal{Y}$ , and  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ , there exists  $\mathcal{C} \in 2^{\mathcal{Y}}$  such that  $v \succ_a v'$  for all  $v' \in \mathcal{C}$  and

<sup>11</sup>If one instead normalizes  $\sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} | \mathbf{y}) = 1$ , then we get the consideration rule of [Aguiar \(2017\)](#).

$$\eta_a(\mathcal{C} \cup \{v\} \mid \mathbf{y}) > 0.$$

(A3') For each  $a \in \mathcal{A}$ ,  $\mathcal{C} \in 2^{\mathcal{Y}}$ , and  $\mathbf{y}, \mathbf{y}^* \in \bar{\mathcal{Y}}^A$ , such that  $\mathbf{y}$  is different from  $\mathbf{y}^*$  just in one component  $a^*$ ,

$$(i) \ a^* \notin \mathcal{N}_a \text{ or } y_{a^*}, y_{a^*}^* \notin \mathcal{C} \implies \eta_a(\mathcal{C} \mid \mathbf{y}) = \eta_a(\mathcal{C} \mid \mathbf{y}^*);$$

$$(ii) \ a^* \in \mathcal{N}_a, y_{a^*} \in \mathcal{C} \text{ and } y_{a^*}^* \notin \mathcal{C} \implies \eta_a(\mathcal{C} \mid \mathbf{y}) > \eta_a(\mathcal{C} \mid \mathbf{y}^*).$$

Assumption A3'(i) states that the attention a person pays to a given set is invariant to the choices of those who are not connected with the person and to alternatives that do not enter the set. Assumption A3'(ii) means that switches of friends to a new alternative boost the attention for all sets that contain this new alternative. Note that Assumption A3' does not assume that different friends affect consideration probabilities symmetrically. That is, the model allows the possibility that some friends have a bigger effect than others. Since the probability of facing consideration set  $\mathcal{C}$  is proportional to the inverse of the total attention,  $\sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} \mid \mathbf{y})$ , even if peers are switching between alternatives that are not elements of  $\mathcal{C}$ , the probability of facing  $\mathcal{C}$  may change.

The next proposition states that, with this general alternative specification, the network structure and the profile of preferences can be uniquely recovered from the conditional choice probabilities. Though the attention mechanism is just partially-identified, it is point identified under additional restrictions that reduce the dimensionality of the problem; we describe some of these restrictions in the next result.

**Proposition 5.3.** *Suppose that Assumptions A1', A2, and A3' are satisfied. Then, the set of connections  $\Gamma$  and the profile of strict preferences  $\succ$  are point identified from P. If additionally for any  $a \in \mathcal{A}$ ,  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ , and  $\mathcal{C}, \mathcal{D} \in 2^{\mathcal{Y}}$  one of the following holds*

$$(i) \text{ (Manzini and Mariotti, 2014)} \ \eta_a(\mathcal{C} \mid \mathbf{y}) = \prod_{v \in \mathcal{C}} \eta_a(\{v\} \mid \mathbf{y});$$

$$(ii) \text{ (Dardanoni et al., 2020)} \ |\mathcal{C}| = |\mathcal{D}| \implies \eta_a(\mathcal{C} \mid \mathbf{y}) = \eta_a(\mathcal{D} \mid \mathbf{y});$$

$$(iii) \ \eta_a(\mathcal{C} \mid \mathbf{y}) = \sum_{v \in \mathcal{C}} \eta_a(\{v\} \mid \mathbf{y});$$

$$(iv) \ \eta_a(\mathcal{C} \mid \mathbf{y}) = \eta_a(\{v^*\} \mid \mathbf{y}), \text{ where } v^* \in \mathcal{C} \text{ satisfies } v^* \succ_a v' \text{ for all } v' \in \mathcal{C};$$

then  $\{\eta_a\}_{a \in \mathcal{A}}$  is also pointidentified.

Condition (i) in Proposition 5.3 restates our main result using the consideration set formation model of Manzini and Mariotti (2014) —see also Manski (1977). Condition (ii) shows that the model analyzed in Dardanoni et al. (2020) —where the sets of equal cardinality are considered with equal probability— also imposes enough restrictions to recover consideration probabilities. Conditions (iii) and (iv) are new. Condition (iii) is similar to condition (i), but defines “aggregate” attention as a sum of singleton-attentions. Condition (iv) postulates that consideration probability of a set only depends on the best alternative in that set.

## 6. Estimation of the Model Parameters

This section provides an estimator of the model parameters that uses discrete-time data (Dataset 2) and assesses its quality in Monte Carlo simulations.

Let  $\theta = (\Gamma, \succ, Q)$  be an element of the space of possible parameters we want to estimate. (We normalize the intensity parameter  $\lambda_a$  to 1 for all  $a \in \mathcal{A}$ .) Note that since the number of people in the network and the number of choices is finite, the parameter space for  $\Gamma$  and  $\succ$  is a finite set. For each  $\theta$  we can construct the transition rate matrix  $\mathcal{M}(\theta)$  using Equations (1) and (2). In turn, this information allows us to calculate the transition matrix

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

Given a sample of network configurations  $\{\mathbf{y}_t\}_{t=0}^T$ , we can use the latter to build the log-likelihood function  $L_T(\theta) = \sum_{t=0}^{T-1} \ln \mathcal{P}_{\iota(\mathbf{y}_t), \iota(\mathbf{y}_{t+1})}(\theta, \Delta)$ , where  $\iota(\mathbf{y}) \in \{1, 2, \dots, \bar{\mathcal{Y}}^A\}$  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, the maximum likelihood estimator of the true parameter value can be defined as

$$\hat{\theta}_T = \arg \max_{\theta} L_T(\theta).$$

To evaluate the finite-sample properties of our estimator we conducted several Monte Carlo experiments. In all experiments we simulated discrete-time data from the specification presented in Example 2 in Section 3. In particular, we assumed that

$$2 \succ_1 1, \quad 1 \succ_2 2, \quad 2 \succ_3 1, \quad 1 \succ_4 2, \quad 1 \succ_5 2,$$

$$Q(v | 0) = \frac{1}{4}, \quad Q(v | 1) = \frac{3}{4}, \quad Q(v | 2) = \frac{7}{8},$$

and

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

First, we estimate  $Q$  under the assumption that the network structure and preferences are known. The experiment was replicated 1000 times for 6 different sample sizes. The results of these simulations are presented in Table 5. The estimator of  $Q$  performs well in terms of the mean bias and the root mean squared error. As expected, the bias and the root mean squared error decrease with the sample size. (See Appendix C for more details on how the data was generated.)

Next we estimate the whole parameter vector since our method allows to consistently estimate the network structure and the preference orders of individuals as well. Since in our example, without any restrictions, there are  $2^{A(A-1)} = 1,048,576$  possible networks and  $(Y!)^A = 32$  strict preference orders, to make the problem computationally tractable we restricted the parameter space for  $\Gamma$  by making the following assumptions: (i) each person has at most two friends; (ii) the attention mechanism is invariant across people and alternatives; and (iii) the network is undirected. As a result, the number of possible networks becomes 112 (the number of possible preference orders is

**Table 5** – Bias and Root Mean Squared Error (RMSE) ( $\times 10^{-3}$ )

Attention Probabilities		Sample Size					
		10	50	100	500	1000	5000
Q( $v$   0)	Bias	213.6	164.9	133.2	65.1	42.3	9.4
	RMSE	259.1	172.9	137.4	66.6	43.3	10.2
Q( $v$   1)	Bias	66.4	53	46.9	27.2	18	2.7
	RMSE	103.9	66.9	55.6	31	20.9	5.6
Q( $v$   2)	Bias	53.8	49.8	42.9	23.6	15.1	0.3
	RMSE	94.3	64.6	52.8	27.2	17.8	4.4

Notes: The sample sizes of 10, 50, 100, 500, 1000, and 5000 correspond to  $\Delta$  equal to 2500, 500, 250, 50, 25, and 5. The number of replications is 1000.

still 32). The experiment was conducted 500 times for different sample sizes. Table 6 presents the results of these simulations. With just 50 observations the network structure is correctly estimated 94.4 percent times. For the sample size of 500 the network structure and the preferences are correctly estimated in all simulations.

**Table 6** – Correctly Estimated Network & Preferences

Sample Size	10	50	100	500
Network	32.4%	94.4%	99.8%	100%
Preferences	34.6%	85.6%	97.6%	100%
Network & Preferences	13.4%	83%	97.4%	100%

Notes: The sample sizes of 10, 50, 100, and 500 correspond to  $\Delta$  equal to 2500, 500, 250, and 50. The number of replications is 500.

## 7. Empirical Application

In this section, we apply our framework to an experimental dataset used by Bai et al. (2019) to test models that aim to predict the visual focus of attention of people. In each experiment, a group of people were asked to play the Resistance game. This is a party game in which people communicate with each other in order to find the deceptive players.<sup>12</sup> The game was implemented in several rounds and lasted about 30 minutes. For the purpose of our analysis, it is important to remark that all participants were seated around a circle. A tablet was placed in front of each player and it recorded the direction of the player’s sight every 1/3 of a second. The structural approach we offer allows us to disentangle and estimate two key forces in explaining the data: directional sight preferences of players and peer effects in gazing.

<sup>12</sup>See [https://en.wikipedia.org/wiki/The\\_Resistance\\_\(game\)](https://en.wikipedia.org/wiki/The_Resistance_(game)) for a more detailed description of the game.

The two determinants of visual focus of attention that we want to recover for each person are motivated by two well-known observations: First, visual designers create online platforms under the premise that people scan certain areas of the screen before others. In particular, it is believed that people spend more time looking to the left side of the screen as compared to the right side.<sup>13</sup> The so-called left-to-right bias has been also documented by the experimental studies of Spalek and Hammad (2004, 2005), and Reutskaja et al. (2011).<sup>14</sup> Second, in social environments, it has been observed that people automatically redirect their visual attention by following others’ gaze orientation, a phenomenon called gaze following.<sup>15</sup> We believe our model is particularly well-suited to separate these two forces from the observed repeated choices.

Before presenting the main results, we describe the dataset we use and the main assumptions we invoke in the application.

**Data** The dataset contains five experiments of up to eight people. For our study, we use the data coming from the two experiments that have five players. We refer to them as Experiments 1 and 2. Experiments 1 and 2 contain 4248 and 7875 observations, respectively. Although the frequency of sight measure seems to be rather high (3 measurements per second), 19 and 26 percent of the observations in Experiments 1 and 2, respectively, have more than one player changing the direction of sight between consecutive measurements. Thus, we treat the datasets generated by the two experiments as discrete time data (i.e., Dataset 2 in Section 4).

**Choice sets** To capture possible preferences for direction in visual focus of attention, we aggregate the data to three options: left, right, and tablet.<sup>16</sup> (The original dataset has five options as each person can look at one of her four opponents or her tablet.) As a result, every measurement indicates the direction of sight of every player at the moment of recording (e.g., Player 1 looks to the left). Formally, from the initial dataset we construct a new dataset on the joint configuration of choices  $\mathbf{y} = (y_a)_{a \in \mathcal{A}} \in \mathcal{Y}^{\mathcal{A}} = \times_{a \in \mathcal{A}} \{t_a, l_a, r_a\}$ , where:  $y_a = t_a$  if, at the moment of recording, Player  $a$  looks at her own tablet;  $y_a = l_a$  if Player  $a$  looks to her left; and  $y_a = r_a$  if Player  $a$  looks to her right. Players were seated clockwise. Thus, the left side for Players  $a = 1, 2, 3$ , is defined as  $t_a = \{a + 1, a + 2\}$  and for Players 4 and 5 it is  $t_4 = \{5, 1\}$  and  $t_5 = \{1, 2\}$ , respectively. Similarly, the right side is defined as  $r_a = \{a - 1, a - 2\}$  for Players  $a = 3, 4, 5$ , and as  $r_2 = \{1, 5\}$  and  $r_1 = \{5, 4\}$  for Players 2 and 1, respectively. Note that the choice sets differ across players.<sup>17</sup> This feature can be incorporated in our model, as none of our results crucially depend on it.

**Consideration probabilities and choices of friends** We assume that consideration probabilities on the direction of sight (i.e., left, right, or tablet) vary with the number of peers that look in the

<sup>13</sup>See, for example, <https://www.nngroup.com/articles/horizontal-attention-original-research/>.

<sup>14</sup>See also Maass and Russo (2003) and reference therein.

<sup>15</sup>See, for example, Gallup et al. (2012) and <https://www.nationalgeographic.com/science/article/what-are-you-looking-at-people-follow-each-others-gazes-but-without-a-tipping-point#close.21>.

<sup>16</sup>This aggregation has several other advantages. For instance, the induced model has more players than alternatives. It also reduces the cardinality of the outcome space and the possible number of preference orders ( $3^5 = 243$  vs.  $5^5 = 3125$ ).

<sup>17</sup>For instance, if Players 1 and 4 look to the left, they look at different sets of players: Player 1 looks at either Player 2 or Player 3, Player 4, however, looks at either Player 5 or Player 1.

same direction. That is, each player is more likely to consider looking at a particular area if other players are doing so. For each Player  $a$  we say Player  $a'$  is looking in the direction  $y_a$  if  $y_a \cap y_{a'} \neq \emptyset$ . For example, suppose that (at the moment of deciding where to look at) Player 1 observes that  $y_2 = t_2$ ,  $y_3 = r_3$ ,  $y_4 = l_4$ , and  $y_5 = l_5$ . That is, 3 players are looking in the direction of Player 1 (Players 3, 4, and 5), 3 players are looking in the direction of Player 2 (Players 2, 3, and 5), only Player 4 is looking in the direction of Player 5, and no one is looking in the direction of Players 3 and 4. As a result, we state that 3 people are looking to the “left” side of Player 1 ( $l_1 \cap y_{a'} \neq \emptyset$  only for  $a' = 2, 3, 5$ ), 1 person is looking to the “right” side of Player 1 ( $r_1 \cap y_{a'} \neq \emptyset$  only for  $a' = 4$ ), and 3 people are looking at Player 1 ( $t_1 \cap y_{a'} \neq \emptyset$  for  $a' = 3, 4, 5$ ). In this situation, the probability that Player 1 considers left is  $Q_1(l_1 | 3)$ , right is  $Q_1(r_1 | 1)$ , and tablet is  $Q_1(t_1 | 3)$ .

**Network structure** Given the small number of players and the fact that participants are playing a party game with monetary prizes, we think it is natural to assume the network is complete. That is, we assume each player is connected to all other players in the group.

**Preferences and the default option** Since, in this setting, there is no reason for us to treat one of the directions as the default option (i.e., always considered and the least preferred), we use the specification of the model with no default option presented in Section 5.2 for the estimates. Also, we do not impose any restrictions on preferences neither for each person nor across different people. Thus, in total, we have 3 possible preference orders for each player and  $3^5 = 243$  possible combinations of preference orders for the five players.

Before presenting the main estimates, let us display some raw data description. Using the data from the experiments we first construct marginal shares of each option for every player in Experiments 1 and 2. This information is displayed in Tables 7 and 8 below.

**Table 7** – Marginal Shares in the Experiment 1 (%)

	Player 1	Player 2	Player 3	Player 4	Player 5
Left	24.58	50.66	56.57	78.32	55.25
Tablet	6.97	2.71	0.05	2.68	0.94
Right	68.45	46.63	43.38	19	43.81

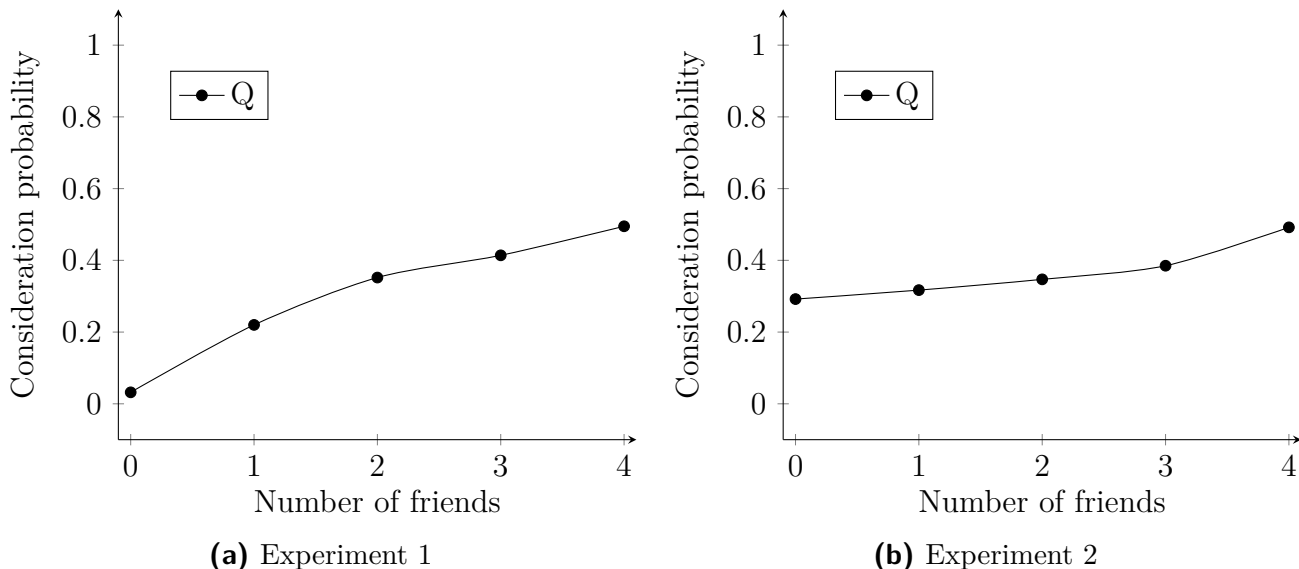
Notes: The sample size is 4248.

Looking at Table 7 we see that, in Experiment 1, all the players spend very little time looking at the tablet. In addition, all players but Player 1 look to the left more often than to the right. Table 8 displays rather different results for Experiment 2. In particular, the second experiment has more balanced shares. Also, looking at own tablet is not longer the least selected option for all the players. For example, Player 2 looks to her tablet more often than to the left or to the right. In addition, three out of the five players look to the right more often than to the left. Based on these initial results, one may expect considerable differences in the two experiments regarding preferences for directional sight. Interestingly, we get that this is not the case.

**Table 8** – Marginal Shares in the Experiment 2 (%)

	Player 1	Player 2	Player 3	Player 4	Player 5
Left	39.7	34.9	30.08	28.98	50.16
Tablet	20.39	37.82	4.88	14.59	16.22
Right	39.91	27.28	65.04	56.43	33.62

Notes: The sample size is 7875.



**Figure 1** – Consideration probability as a function of the number of friends looking in the same direction. Model I(a).

We estimate a few specifications of the model. We present two of them next and display the other ones in Appendix D.2. In Model I(a), we assume there is no heterogeneity in consideration probabilities across options and players, i.e.,  $Q(\cdot) = Q_a(v | \cdot)$  for all  $a$  and  $v$ . Model I(b) adds heterogeneity in consideration probabilities across players. All the models we estimate allow for unrestricted heterogeneity in preferences regarding directional visual sight.

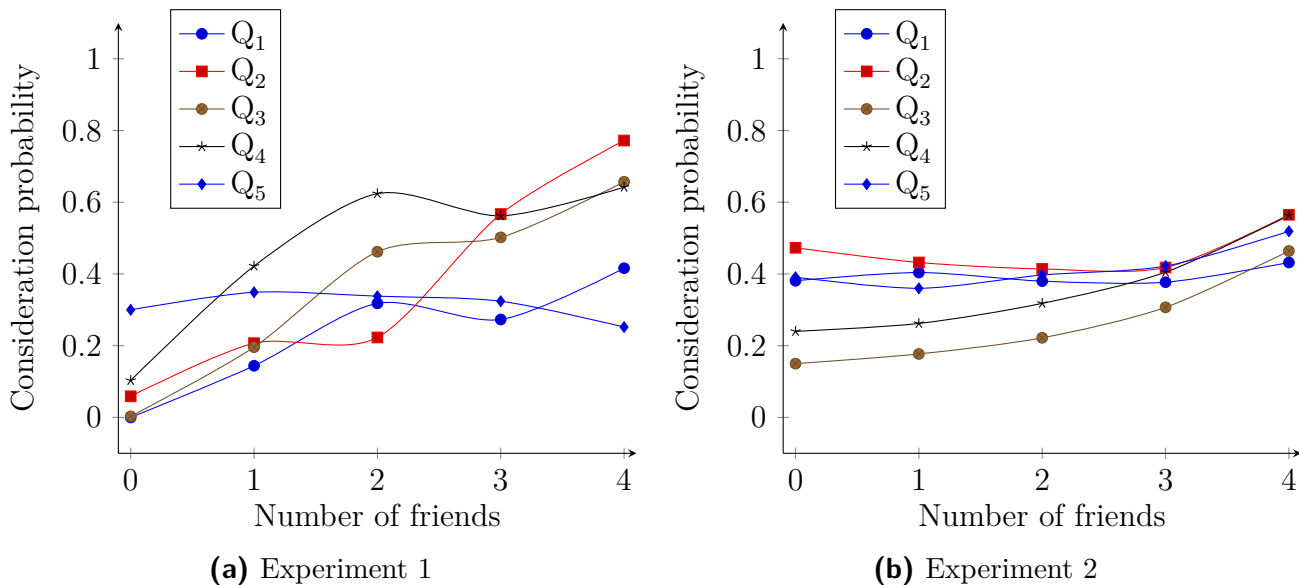
Figure 1 shows the estimates for the consideration probabilities of Model I(a) for the two experiments. Note that although monotonicity of  $Q$  in the number of friends was not imposed in the estimation, the estimated probabilities are indeed monotone in Experiment 2 and are very close to be monotone in Experiment 1. The estimated preferences for directional sight coincide for all players in the two experiments. Specifically, all players prefer looking to the left, then to tablet, and then to the right. That is, we obtain that, for each  $a \in \mathcal{A}$ ,

$$l_a \succ_a t_a \succ_a r_a.$$

Thus, in the first specification of our model, all the players (in Experiments 1 and 2) show the left-to-right bias.

Figure 2 shows the estimates for the consideration probabilities for Model I(b) in Experiments 1 and 2. Similar to the case with homogeneity, many of the consideration probabilities for each





**Figure 2** – Consideration probabilities for different players as functions of the number of friends looking in the same direction. Model I(b).

direction of sight are indeed increasing in the number of players looking at that direction. The estimated preference orders coincide in Models I(a) and (b), for the two experiments. That is, here again, all the players prefer looking to the left, then to tablet, and then to the right.

To sum up, despite considerable differences in raw marginal shares (see Tables 7 and 8), the preferences for directional sight of all players in the two experiments coincide. Specifically, all players prefer to look from left to right after controlling for unobserved consideration sets that are determined by visual focus of attention of others. As a robustness check, in order to address possible concerns with the dynamic of players’ interactions in different stages of the game, we estimated the previous two models using the last half of observations and the middle half of observations (i.e., we did not use the first and the last quarter of observations). In both cases the estimated consideration probabilities are similar to the ones estimated using the whole sample, and the estimated preferences of all players are again  $l_a \succ_a t_a \succ_a r_a$  (see Appendix D.1 for further details). As we mentioned earlier, this type of preferences are consistent with the left-to-right bias invoked by online platform designers that has been also documented by experimental work.

## 8. Final Remarks

This paper adds peer effects to the consideration set models. It does so by combining the dynamic model of social interactions of Blume (1993, 1995) with the (single-agent) model of random consideration sets of Manski (1977) and Manzini and Mariotti (2014). The model we build differs from most of the social interaction models in that the choices of friends do not affect preferences

but the subset of options that people end up considering. In our model, the network structure affects the nature and strength of the mistakes that people make. Thus, we can think of our model as a model of peer effects in mistakes.

From an applied perspective, changes in the choices of friends induce stochastic variation of the considerations sets. We show that this variation can be used to recover the main parts of the model. On top of nonparametrically recovering the preference ranking of each person and the attention mechanism, we identify the set of connections or edges between the people in the network. The identification strategy allows unrestricted heterogeneity across people.

We propose a consistent estimator of model parameters and apply it to an experimental dataset. The structural approach we offer allows us to identify and estimate two main determinants on people visual focus of attention: directional sight preferences and peer effects in gazing. Our results are consistent with the existing literature. We believe our approach to peer effects in consideration sets could be incorporated in various empirical studies of technology adoption and diffusion. For instance, Aral et al. (2009) develop a dynamic matched sample estimation framework to distinguish influence and homophily effects in the day-by-day adoption of a mobile service application (i.e., Yahoo! Go). A similar dynamic matched sample estimation framework could be used to recover preferences over adoption and the influence of neighbor adoption rates on likelihood of considering the possibility of incorporating the mobile service application.

## References

- Abaluck, Jason and Abi Adams (2021) “What do consumers consider before they choose? Identification from asymmetric demand responses,” *The Quarterly Journal of Economics*, Accepted.
- Aguiar, Victor H (2017) “Random categorization and bounded rationality,” *Economics Letters*, 159, 46–52.
- Aguiar, Victor H., Maria Jose Boccardi, and Mark Dean (2016) “Satisficing and stochastic choice,” *Journal of Economic Theory*, 166, 445 – 482, <https://doi.org/10.1016/j.jet.2016.08.008>.
- Aguiar, Victor H, Maria Jose Boccardi, Nail Kashaev, and Jeongbin Kim (2021) “Random utility and limited consideration.”
- Aguiar, Victor H (c) Nail Kashaev (2021) “Identification and Estimation of Discrete Choice Models with Unobserved Choice Sets.”
- Allen, Beth (1982) “Some stochastic processes of interdependent demand and technological diffusion of an innovation exhibiting externalities among adopters,” *International Economic Review*, 23 (3), 595–608.

- Aral, Sinan, Lev Muchnik, and Arun Sundararajan (2009) “Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks,” *Proceedings of the National Academy of Sciences*, 106 (51), 21544–21549.
- Arnold, Barry C and S James Press (1989) “Compatible conditional distributions,” *Journal of the American Statistical Association*, 84 (405), 152–156.
- Bai, Chongyang, Srijan Kumar, Jure Leskovec, Miriam Metzger, Jay F Nunamaker Jr, and VS Subrahmanian (2019) “Predicting the Visual Focus of Attention in Multi-Person Discussion Videos.,” in *IJCAI*, 4504–4510.
- Barseghyan, Levon, Maura Coughlin, Francesca Molinari, and Joshua C Teitelbaum (2021a) “Heterogeneous choice sets and preferences,” *Econometrica*, forthcoming.
- Barseghyan, Levon, Francesca Molinari, and Matthew Thirkettle (2021b) “Discrete choice under risk with limited consideration,” *American Economic Review*, forthcoming.
- Blevins, Jason R (2017) “Identifying restrictions for finite parameter continuous time models with discrete time data,” *Econometric Theory*, 33 (3), 739–754.
- (2018) “Identification and estimation of continuous time dynamic discrete choice games,” Technical report, Mimeo, Ohio State University.
- Blume, Lawrence E (1993) “The statistical mechanics of strategic interaction,” *Games and economic behavior*, 5 (3), 387–424.
- (1995) “The statistical mechanics of best-response strategy revision,” *Games and Economic Behavior*, 11 (2), 111–145.
- Blume, Lawrence E, William A Brock, Steven N Durlauf, and Yannis M Ioannides (2011) “Identification of social interactions,” in *Handbook of social economics*, 1, 853–964: Elsevier.
- Blume, Lawrence E, William A Brock, Steven N Durlauf, and Rajshri Jayaraman (2015) “Linear social interactions models,” *Journal of Political Economy*, 123 (2), 444–496.
- Blume, Lawrence E and Steven Durlauf (2003) “Equilibrium concepts for social interaction models,” *International Game Theory Review*, 5 (3), 193–209.
- Bonaldi, Pietro, Ali Hortaçsu, and Jakub 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.
- Borah, Abhinash and Christopher Kops (2018) “Choice via Social Influence,” Technical report, Mimeo.
- Brady, Richard L and John Rehbeck (2016) “Menu-Dependent Stochastic Feasibility,” *Econometrica*, 84 (3), 1203–1223.

- Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2020) “Peer effects in networks: A survey,” *Annual Review of Economics*, 12, 603–629.
- Brock, William A and Steven N Durlauf (2001) “Discrete choice with social interactions,” *The Review of Economic Studies*, 68 (2), 235–260.
- (2002) “A multinomial-choice model of neighborhood effects,” *American Economic Review*, 92 (2), 298–303.
- Caeyers, Bet et al. (2014) “Peer effects in development programme awareness of vulnerable groups in rural Tanzania,” Technical report, Centre for the Study of African Economies, University of Oxford.
- Caplin, Andrew, Mark Dean, and John Leahy (2019) “Rational inattention, optimal consideration sets, and stochastic choice,” *The Review of Economic Studies*, 86 (3), 1061–1094.
- Cattaneo, Matias D, Xinwei Ma, Yusufcan Masatlioglu, and Elchin Suleymanov (2020) “A random attention model,” *Journal of Political Economy*, 128 (7), 2796–2836.
- Chambers, Christopher P, Tugce Cuhadaroglu, and Yusufcan Masatlioglu (2019) “A Random Attention Model,” *Mimeo*.
- Conlon, Christopher T and Julie Holland Mortimer (2013) “Demand estimation under incomplete product availability,” *American Economic Journal: Microeconomics*, 5 (4), 1–30.
- Crawford, Gregory S, Rachel Griffith, and Alessandro Iaria (2021) “A survey of preference estimation with unobserved choice set heterogeneity,” *Journal of Econometrics*, 222 (1), 4–43.
- Dardanoni, Valentino, Paola Manzini, Marco Mariotti, and Christopher J Tyson (2020) “Inferring cognitive heterogeneity from aggregate choices,” *Econometrica*, 88 (3), 1269–1296.
- De Paula, Aureo (2017) “Econometrics of network models,” in *Advances in Economics and Econometrics: Theory and Applications, Eleventh World Congress*, 268–323, Cambridge University Press Cambridge.
- De Paula, Áureo, Imran Rasul, and Pedro Souza (2018) “Recovering social networks from panel data: identification, simulations and an application.”
- Draganska, Michaela and Daniel Klapper (2011) “Choice set heterogeneity and the role of advertising: An analysis with micro and macro data,” *Journal of Marketing Research*, 48 (4), 653–669.
- Duflo, Esther and Emmanuel Saez (2003) “The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment,” *The Quarterly journal of economics*, 118 (3), 815–842.
- Enikolopov, Ruben, Alexey Makarin, and Maria Petrova (2020) “Social media and protest participation: Evidence from Russia,” *Econometrica*, 88 (4), 1479–1514.

- Gallup, Andrew C, Joseph J Hale, David JT Sumpter, Simon Garnier, Alex Kacelnik, John R Krebs, and Iain D Couzin (2012) “Visual attention and the acquisition of information in human crowds,” *Proceedings of the National Academy of Sciences*, 109 (19), 7245–7250.
- Gaynor, Martin, Carol Propper, and Stephan Seiler (2016) “Free to choose? Reform, choice, and consideration sets in the English National Health Service,” *American Economic Review*, 106 (11), 3521–57.
- Godes, David and Dina Mayzlin (2004) “Using online conversations to study word-of-mouth communication,” *Marketing science*, 23 (4), 545–560.
- Goeree, Michelle Sovinsky (2008) “Limited information and advertising in the US personal computer industry,” *Econometrica*, 76 (5), 1017–1074.
- Graham, Bryan S (2015) “Methods of identification in social networks,” *Annu. Rev. Econ.*, 7 (1), 465–485.
- Horan, Sean (2019) “Random consideration and choice: A case study of “default” options,” *Mathematical Social Sciences*, 102, 73–84.
- Kaiser, Mark S and Noel Cressie (2000) “The construction of multivariate distributions from Markov random fields,” *Journal of Multivariate Analysis*, 73 (2), 199–220.
- Lazzati, Natalia (2020) “Co-Diffusion of Technologies in Social Networks,” *American Economic Journal: Microeconomics*, Forthcoming.
- Lleras, Juan Sebastian, Yusufcan Masatlioglu, Daisuke Nakajima, and Erkut Y Ozbay (2017) “When more is less: Limited consideration,” *Journal of Economic Theory*, 170, 70–85.
- Maass, Anne and Aurore Russo (2003) “Directional bias in the mental representation of spatial events: Nature or culture?” *Psychological science*, 14 (4), 296–301.
- Manresa, Elena (2013) “Estimating the structure of social interactions using panel data,” *Unpublished Manuscript. CEMFI, Madrid*.
- Manski, Charles F (1977) “The structure of random utility models,” *Theory and decision*, 8 (3), 229–254.
- Manzini, Paola and Marco Mariotti (2014) “Stochastic choice and consideration sets,” *Econometrica*, 82 (3), 1153–1176.
- Masatlioglu, Yusufcan, Daisuke Nakajima, and Erkut Y. Ozbay (2012) “Revealed Attention,” *American Economic Review*, 102 (5), 2183–2205, [10.1257/aer.102.5.2183](https://doi.org/10.1257/aer.102.5.2183).
- Mehta, Nitin, Surendra Rajiv, and Kannan Srinivasan (2003) “Price uncertainty and consumer search: A structural model of consideration set formation,” *Marketing science*, 22 (1), 58–84.

- Qiu, Liangfei, Zhan Shi, and Andrew B Whinston (2018) “Learning from your friends’ check-ins: An empirical study of location-based social networks,” *Information Systems Research*, 29 (4), 1044–1061.
- Reutskaja, Elena, Rosemarie Nagel, Colin F Camerer, and Antonio Rangel (2011) “Search dynamics in consumer choice under time pressure: An eye-tracking study,” *American Economic Review*, 101 (2), 900–926.
- Roberts, John H and James M Lattin (1991) “Development and testing of a model of consideration set composition,” *Journal of Marketing Research*, 28 (4), 429–440.
- Spalek, Thomas M and Sherief Hammad (2004) “Supporting the attentional momentum view of IOR: Is attention biased to go right?” *Perception & Psychophysics*, 66 (2), 219–233.
- (2005) “The left-to-right bias in inhibition of return is due to the direction of reading,” *Psychological Science*, 16 (1), 15–18.

## A. Proofs

**Proof of Proposition 3.1:** For an irreducible, finite-state, continuous Markov chain the steady-state  $\mu$  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  $a \in \mathcal{A}$ ,  $v \in \mathcal{Y}$ , and  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ , we have that

$$1 > P_a(v | \mathbf{y}) = Q_a(v | y_a, N_a^v(\mathbf{y})) \prod_{v' \in \mathcal{Y}, v' \succ_a v} (1 - Q_a(v' | y_a, N_a^{v'}(\mathbf{y}))) > 0.$$

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. ■

**Proof of Proposition 4.1:** By A1,  $P_a(\cdot | \mathbf{y})$  has full support for each  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ . By A2 and A3,  $P_a(v | \mathbf{y})$  is strictly decreasing in  $N_a^{v'}(\mathbf{y})$  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  $\Gamma = (\mathcal{A}, e)$ . Also, from variation in  $N_a^{v'}(\mathbf{y})$  for each  $v' \neq v$ , we can recover Person  $a$ ’s upper level set that corresponds to option  $v$ . That is,

$$\{v' \in \mathcal{Y} : v' \succ_a v\}$$

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,

$$P_a(y_a^* | \mathbf{y}) = Q_a(y_a^* | y_a, N_a^{y_a^*}(\mathbf{y})).$$

It follows that we can recover  $Q_a(y_a^* | y_a, N_a^{y_a^*}(\mathbf{y}))$  for each  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ . By proceeding in descending preference ordering we can then recover  $Q_a(v | y_a, N_a^v(\mathbf{y}))$  for all  $v \in \mathcal{Y}$  (and each  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ ). ■

**Proof of Proposition 4.2:** Since  $\lim_{\Delta \rightarrow 0} \mathcal{P}(\Delta) = \mathcal{M}$ , we can recover transition rate matrix from the data. Recall that

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

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

$$\sum_{v \in \bar{\mathcal{Y}}} \lambda_a P_a(v | \mathbf{y}) = \lambda_a \sum_{v \in \bar{\mathcal{Y}}} P_a(v | \mathbf{y}) = \lambda_a.$$

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

**Proof of Proposition 4.3:** This proof builds on Theorem 1 of Blevins (2017) and Theorem 3 of Blevins (2018). For the present case, it follows from these two theorems, that the transition rate matrix  $\mathcal{M}$  is generically identified if, in addition to the conditions in Proposition 4.3, we have that

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

This condition is always satisfied if  $A > 1$ . Thus, identification of  $\mathcal{M}$  follows because, by A2,  $A \geq 2$ . We can then uniquely recover  $(P_a)_{a \in \mathcal{A}}$  from  $\mathcal{M}$  as in the proof of Proposition 4.2 ■

**Proof of Proposition 5.1:** Note that expression (3) can be rewritten as follows

$$P_a(v | \mathbf{y}) = \sum_{\mathcal{C} \subseteq \mathcal{Y}} R_a(v | \mathcal{C}) \prod_{v' \in \mathcal{C}} Q_a(v' | y_a, N_a^{v'}(\mathbf{y})) \prod_{v'' \notin \mathcal{C}} (1 - Q_a(v'' | y_a, N_a^{v''}(\mathbf{y}))) = \\ Q_a(v | y_a, N_a^v(\mathbf{y})) \sum_{\mathcal{C} \subseteq \mathcal{Y}, v \in \mathcal{C}} R_a(v | \mathcal{C}) \prod_{v' \in \mathcal{C} \setminus \{v\}} Q_a(v' | y_a, N_a^{v'}(\mathbf{y})) \prod_{v'' \notin \mathcal{C}} (1 - Q_a(v'' | y_a, N_a^{v''}(\mathbf{y}))).$$

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

Let  $\mathbf{y}$  be such that  $N_a^v(\mathbf{y}) = 0$  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

$$N_a^{v'}(\mathbf{y}) = N_a^{v'}(\mathbf{y}') \text{ for all } v' \neq v \text{ and } N_a^v(\mathbf{y}) = 1.$$

Note that

$$P_a(v | \mathbf{y}') / P_a(v | \mathbf{y}) = Q_a(v | y_a, 1) / Q_a(v | y_a, 0).$$

Also

$$P_a(o | \mathbf{y}') / P_a(o | \mathbf{y}) = (1 - Q_a(v | y_a, 1)) / (1 - Q_a(v | y_a, 0)).$$

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

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  $v = 1$ , agent  $a$ , and  $y_a$ . (The proof immediately extends to other agents and alternatives.) We want to recover  $R_a(1 | \mathcal{C})$  for all  $\mathcal{C}$ . To simplify the exposition we will write

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

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

$$P_a(1 | \mathbf{y}) / Q_a(1 | y_a, N_a^1(\mathbf{y})) = \sum_{\mathcal{C} \subseteq \mathcal{Y}, v \in \mathcal{C}} R_a(v | \mathcal{C}) \prod_{k \in \mathcal{C} \setminus \{v\}} Q^1(k | N_a^k(\mathbf{y})) \prod_{k \notin \mathcal{C} \setminus \{v\}} Q^0(k | N_a^k(\mathbf{y})).$$

To present the ideas more clear let  $A(N, Y)$  be the matrix of coefficients in front of the  $R_a$ '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

$$\prod_{k \in \mathcal{Y}} Q^{1(k \in \mathcal{C})}(k | N^k)$$

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

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

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

**Proof.**

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

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

and

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



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 submatrix 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

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

In summary, we have that

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

**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  $\det(A(1, 2)) = Q^1(2 | 1) - Q^1(2 | 0) > 0$ . Also, any  $A(N, 2)$  with  $N \geq 1$  will have full column rank because  $A(1, 2)$  is a submatrix of  $A(N, 2)$  with the same number of columns.

Finally, note that  $A(1, 3)$  does not have 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  $N \geq 1$ ,

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

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

**Proof of Proposition 5.2:** This proof is divided in three steps.

**Step 1. (Identification of the Set of Connections)** Take any two different people with

arbitrary designations  $a_1$  and  $a_2$ . Note that if  $a_2 \notin \mathcal{N}_{a_1}$ , then  $P_{a_1}(v | \mathbf{y}) = P_{a_1}(v | \mathbf{y}')$  for any  $\mathbf{y}$  and  $\mathbf{y}'$  such that  $y_a = y'_a$  for all  $a \neq a_2$  and  $y_{a_1} \neq v$ . Also, let  $v_{a_1}^*$  be the best preferred alternative of  $a_1$ . Then, by A1 and A3, for any  $\mathbf{y}$  such that  $y_{a_1} \neq v_{a_1}^*$

$$P_{a_1}(v_{a_1}^* | \mathbf{y}) = Q_{a_1}(v_{a_1}^* | \mathbf{y})$$

is constant in  $y_{a_2}$  if and only if  $a_2 \notin \mathcal{N}_{a_1}$ . Altogether,  $a_2 \notin \mathcal{N}_{a_1}$  if and only if  $P_{a_1}(v | \mathbf{y})$  with  $y_{a_1} \neq v$  is constant in  $y_{a_2}$ . As a result, we can identify whether  $a_2$  is in the set of friends of  $a_1$ . Since the choice of  $a_1$  and  $a_2$  was arbitrary, we can identify the set of connections  $\Gamma$ . Note that for this result we only need  $Y \geq 2$ .

**Step 2. (Identification of the Preferences)** Fix some Person  $a_1$ . We will show the result for a set of alternatives  $\mathcal{Y} = \{1, 2, 3\}$  of size 3. (The proof easily extends to the case of more alternatives. The only cost is extra notation.) For arbitrary designation of the three options  $v_1, v_2, v_3$ , and any  $a_2 \in \mathcal{N}_{a_1}$  (by A2,  $\mathcal{N}_{a_1} \neq \emptyset$ ) let

$$\Delta P(v_1, v_2, v_3) = P_{a_1}(v_1 | (v_3, \dots, v_3, v_2, v_3, \dots, v_3)) - P_{a_1}(v_1 | (v_3, \dots, v_3, v_3, v_3, \dots, v_3)),$$

where Person  $a_2$  switches from  $v_2$  to  $v_3$ . Define  $\text{sign}(v_1, v_2, v_3) \in \{-, +, 0\}$  be such that

$$\text{sign}(v_1, v_2, v_3) = \begin{cases} - & \text{if } \Delta P(v_1, v_2, v_3) < 0 \\ + & \text{if } \Delta P(v_1, v_2, v_3) > 0 \\ 0 & \text{if } \Delta P(v_1, v_2, v_3) = 0 \end{cases}.$$

Note that  $\text{sign}(\cdot)$  can be computed from the data. Different preference orders over  $\mathcal{Y}$  will imply potentially different values for  $\text{sign}(\cdot)$ . Let  $N_{a_1}$  is the the cardinality of  $\mathcal{N}_{a_1}$ . For example, if  $1 \succ_{a_1} 2 \succ_{a_1} 3$ , then we have that

$$\begin{aligned} \Delta P(1, 2, 3) &= Q_{a_1}(1 | 3, 0) - Q_{a_1}(1 | 3, 0) = 0 \\ \Delta P(1, 3, 2) &= Q_{a_1}(1 | 2, 0) - Q_{a_1}(1 | 3, 0) = 0 \\ \Delta P(3, 1, 2) &= Q_{a_1}(3 | 2, 0)(1 - Q_{a_1}(1 | 2, 1))(1 - Q_{a_1}(2 | 2, N_{a_1} - 1)) \\ &\quad - Q_{a_1}(3 | 2, 0)(1 - Q_{a_1}(1 | 2, 0))(1 - Q_{a_1}(2 | 2, N_{a_1})) \\ \Delta P(3, 2, 1) &= Q_{a_1}(3 | 1, 0)(1 - Q_{a_1}(1 | 1, N_{a_1} - 1))(1 - Q_{a_1}(2 | 1, 1)) \\ &\quad - Q_{a_1}(3 | 1, 0)(1 - Q_{a_1}(1 | 1, N_{a_1}))(1 - Q_{a_1}(2 | 1, 0)), \end{aligned}$$

Note that for  $\Delta P(3, 1, 2)$  and  $\Delta P(3, 2, 1)$  our monotonicity restriction is not informative enough:  $\Delta P(3, 1, 2)$  and  $\Delta P(3, 2, 1)$  can be positive, negative, or equal to zero. The following table displays the values  $\text{sign}(\cdot)$  takes depending on the underlying preference order  $\succ_{a_1}$  for all distinct  $v_1, v_2$ , and

$v_3$ . If the sign of  $\Delta P(v_1, v_2, v_3)$  is not uniquely determined by a strict preference, then we write  $\sim$ .

order/ $(v_1, v_2, v_3)$	(1, 2, 3)	(1, 3, 2)	(2, 1, 3)	(2, 3, 1)	(3, 1, 2)	(3, 2, 1)
$1 \succ_{a_1} 2 \succ_{a_1} 3$	0	0	-	+	$\sim$	$\sim$
$1 \succ_{a_1} 3 \succ_{a_1} 2$	0	0	$\sim$	$\sim$	-	+
$2 \succ_{a_1} 1 \succ_{a_1} 3$	-	+	0	0	$\sim$	$\sim$
$2 \succ_{a_1} 3 \succ_{a_1} 1$	$\sim$	$\sim$	0	0	+	-
$3 \succ_{a_1} 1 \succ_{a_1} 2$	+	-	$\sim$	$\sim$	0	0
$3 \succ_{a_1} 2 \succ_{a_1} 1$	$\sim$	$\sim$	+	-	0	0

Note that we can always distinguish  $1 \succ_{a_1} 2 \succ_{a_1} 3$  from say  $3 \succ_{a_1} 2 \succ_{a_1} 1$  since  $\text{sign}(2, 1, 3)$  is determined for both preference orders and gives different predictions. The only pairs of preference orders that may be observationally equivalent are the following three pairs: (i)  $1 \succ_{a_1} 2 \succ_{a_1} 3$  and  $1 \succ_{a_1} 3 \succ_{a_1} 2$ ; (ii)  $2 \succ_{a_1} 1 \succ_{a_1} 3$  and  $2 \succ_{a_1} 3 \succ_{a_1} 1$ ; and (iii)  $3 \succ_{a_1} 1 \succ_{a_1} 2$  and  $3 \succ_{a_1} 2 \succ_{a_1} 1$ . Although we cannot uniquely identify the order, we can uniquely identify the most preferred alternative. For example, if we know that  $1 \succ_{a_1} 2 \succ_{a_1} 3$  or  $1 \succ_{a_1} 2 \succ_{a_1} 3$  has generated the data, then 1 is the most preferred alternative for agent  $a_1$ .

Assume, without loss of generality, that 3 is the most preferred alternative. Then we can identify  $Q_{a_1}(3 | y_{a_1}, \cdot)$  for any  $y_{a_1} \neq 3$  since

$$Q_{a_1}(3 | y_{a_1}, N_{a_1}^3(\mathbf{y})) = P_{a_1}(3 | \mathbf{y}).$$

Hence, for  $\mathbf{y}$  with  $y_{a_1} = 1$  we have that

$$\frac{P_{a_1}(2 | \mathbf{y})}{1 - P_{a_1}(3 | \mathbf{y})} = \begin{cases} Q_a(2 | 1, N_{a_1}^2(\mathbf{y})), & \text{if } 2 \succ_{a_1} 1 \\ Q_a(2 | 1, N_{a_1}^2(\mathbf{y}))(1 - Q_a(1 | 1, N_{a_1}^1(\mathbf{y}))), & \text{if } 1 \succ_{a_1} 2. \end{cases}$$

Start with a  $\mathbf{y}$  such that  $y_a = 3$  for all  $a \neq a_1$  (and  $y_{a_1} = 1$ ). Consider then changing the  $y_{a_2}$  from 3 to 1 for some  $a_2 \in \mathcal{N}_{a_1}$ . Then, by A3,  $1 \succ_{a_1} 2$  if and only if  $P_{a_1}(2 | \mathbf{y}) / (1 - P_{a_1}(3 | \mathbf{y}))$  strictly decreases in the data.

**Step 3. (Identification of the Attention Mechanism)** Fix some  $a_1$  and let  $y_{a_1}^*$  be the most preferred alternative of Person  $a_1$ . Then we can identify  $Q_{a_1}(y_{a_1}^* | y_{a_1}, N_{a_1}^{y_{a_1}^*}(\mathbf{y}))$  for any  $\mathbf{y}$  such that  $y_{a_1}^* \neq y_{a_1}$  since

$$Q_{a_1}(y_{a_1}^* | y_{a_1}, N_{a_1}^{y_{a_1}^*}(\mathbf{y})) = P_{a_1}(y_{a_1}^* | \mathbf{y}).$$

By proceeding in decreasing preference order we can recover  $Q_{a_1}(y'_{a_1} | y_{a_1}, N_{a_1}^{y'_{a_1}}(\mathbf{y}))$  for any  $y'_{a_1}$  and  $\mathbf{y}$  such that  $y'_{a_1} \neq y_{a_1}$ . Moreover, we can identify

$$\prod_{v' \neq y_{a_1}} (1 - Q_a(v' | y_{a_1}, N_{a_1}^{v'}(\mathbf{y})))$$

Next note that for any  $\mathbf{y}$  such that  $y_{a_1}^* = y_{a_1}$

$$Q_{a_1}(y_{a_1}^* | y_{a_1}, N_{a_1}^{y_{a_1}^*}(\mathbf{y})) = \frac{P_{a_1}(y_{a_1}^* | \mathbf{y}) - \prod_{v' \neq y_{a_1}^*} (1 - Q_{a_1}(v' | y_{a_1}^*, N_{a_1}^{v'}(\mathbf{y})))}{1 - \prod_{v' \neq y_{a_1}^*} (1 - Q_{a_1}(v' | y_{a_1}^*, N_{a_1}^{v'}(\mathbf{y})))}.$$

Hence, we can identify  $Q_{a_1}(y_{a_1}^* | y_{a_1}, N_{a_1}^{y_{a_1}^*}(\mathbf{y}))$  for all  $\mathbf{y}$ . Let  $y_{a_1}^{**}$  be the second best alternative of Person  $a_1$ , then for any  $\mathbf{y}$  such that  $y_{a_1}^{**} = y_{a_1}$  similarly to the case with  $y_{a_1}^*$  we can identify  $Q_{a_1}(y_{a_1}^{**} | y_{a_1}, N_{a_1}^{y_{a_1}^{**}}(\mathbf{y}))$  since

$$Q_{a_1}(y_{a_1}^{**} | y_{a_1}, N_{a_1}^{y_{a_1}^{**}}(\mathbf{y})) = \frac{P_{a_1}(y_{a_1}^{**} | \mathbf{y}) - \prod_{v' \neq y_{a_1}^{**}} (1 - Q_{a_1}(v' | y_{a_1}^{**}, N_{a_1}^{v'}(\mathbf{y})))}{1 - Q_{a_1}(y_{a_1}^* | y_{a_1}, N_{a_1}^{y_{a_1}^*}(\mathbf{y})) - \prod_{v' \neq y_{a_1}^{**}} (1 - Q_{a_1}(v' | y_{a_1}^{**}, N_{a_1}^{v'}(\mathbf{y})))},$$

and thus we recover  $Q_{a_1}(y_{a_1}^{**} | y_{a_1}, N_{a_1}^{y_{a_1}^{**}}(\mathbf{y}))$  for all  $\mathbf{y}$ . By proceeding in decreasing preference order we can recover  $Q_{a_1}(y'_{a_1} | y_{a_1}, N_{a_1}^{y'_{a_1}}(\mathbf{y}))$  for any  $y'_{a_1}$  and  $\mathbf{y}$ . Since the choice of  $a_1$  was arbitrary we can identify  $(Q_a)_{a \in \mathcal{A}}$ .

**Proof of Proposition 5.3: Step 1. (Identification of the Set of Connections)** Take any two different agents  $a_1$  and  $a_2$ . Note that if  $a_2 \notin \mathcal{N}_{a_1}$ , then  $P_{a_1}(v | \mathbf{y}) = P_{a_1}(v | \mathbf{y}')$  for any  $\mathbf{y}$  and  $\mathbf{y}'$  such that  $y_a = y'_a$  for all  $a \neq a_2$  and  $y_{a_1} \neq v$ . Also, if  $v_{a_1}^*$  is the best preferred alternative of  $a_1$ , then by Assumptions A1' and A3', for any  $\mathbf{y}$

$$\frac{P_{a_1}(v_{a_1}^* | \mathbf{y})}{P_{a_1}(o | \mathbf{y})} = \sum_{\mathcal{C} \in 2^{\mathcal{Y}}: v_{a_1}^* \in \mathcal{C}} \eta_{a_1}(\mathcal{C} | \mathbf{y})$$

is constant in choices of Person  $a_2$  if and only if  $a_2 \notin \mathcal{N}_{a_1}$ . Hence,  $a_2 \notin \mathcal{N}_{a_1}$  if and only if  $P_{a_1}(v | \mathbf{y}) / P_{a_1}(o | \mathbf{y})$  is constant in  $y_{a_2}$ . As a result, we can identify whether  $a_2$  is a friend of  $a_1$ . Since the choice of  $a_1$  and  $a_2$  was arbitrary we can identify the whole  $\Gamma$ . Note that for this result to hold we only need  $|\mathcal{Y}| \geq 2$ .

**Step 2. (Identification of the Preferences)** Fix Person  $a_1$  and  $\mathbf{y}^* = (o, o, \dots, o)'$ . Note that

$$\frac{P_{a_1}(v_{a_1}^* | \mathbf{y}^*)}{P_{a_1}(o | \mathbf{y}^*)} = \sum_{\mathcal{C} \in 2^{\mathcal{Y}}: v_{a_1}^* \in \mathcal{C}} \eta_{a_1}(\mathcal{C} | \mathbf{y}^*)$$

will increase if any of friends of  $a_1$  switches to anything else. Let  $v_{a_1}^{**}$  be the second best preferred alternative of  $a_1$ . Then

$$\frac{P_{a_1}(v_{a_1}^{**} | \mathbf{y}^*)}{P_{a_1}(o | \mathbf{y}^*)} = \sum_{\mathcal{C} \in 2^{\mathcal{Y}}: v_{a_1}^* \notin \mathcal{C}, v_{a_1}^{**} \in \mathcal{C}} \eta_{a_1}(\mathcal{C} | \mathbf{y}^*)$$

will increase if any of friends of  $a_1$  switches to anything else but  $v_{a_1}^*$ . Moreover, this probability will not change if any of friends of  $a_1$  switches to  $v_{a_1}^*$ . Hence, we can identify  $v_{a_1}^*$ . Applying the above step in decreasing order, we can identify the whole preference order of Person  $a_1$ . Since the choice

of  $a_1$  was arbitrary we identify preferences of all persons.

**Step 3. (Identification of the Attention Mechanism)** Take any Person  $a_1$  and configuration  $\mathbf{y}$ . Since preferences are identified, assume without loss of generality that  $Y \succ_{a_1} Y-1 \succ_{a_1} Y-2 \succ_{a_1} \dots \succ_{a_1} 2 \succ_{a_1} 1$ . Note that we have the following system of  $Y$  equations

$$\frac{P_{a_1}(k | \mathbf{y})}{P_{a_1}(o | \mathbf{y})} = \begin{cases} \eta_{a_1}(\{1\} | \mathbf{y}), & k = 1, \\ \sum_{\mathcal{C} \subseteq \{1, \dots, k-1\}} \eta_{a_1}(\mathcal{C} \cup \{k\} | \mathbf{y}), & k = 2, \dots, Y. \end{cases}$$

Note that  $\eta_{a_1}$  could have generated the data if and only if it solves the above system of equations. Since, there are  $2^Y - 1$  unknown parameters (recall that attention to the empty set is normalized to be 1) and  $Y$  equations, and there is no single attention parameter that enters more than one equation,  $\eta_{a_1}$  can not be identified without more restrictions. Suppose  $\eta_{a_1}$  is multiplicative. So we only need to identify  $\eta_{a_1}(\{k\} | \mathbf{y})$  for all  $k \in \mathcal{Y}$ . Then, first, we can identify  $\eta_{a_1}(\{1\} | \mathbf{y})$  since

$$\frac{P_{a_1}(1 | \mathbf{y})}{P_{a_1}(o | \mathbf{y})} = \eta_{a_1}(\{1\} | \mathbf{y}).$$

Next,

$$\frac{P_{a_1}(2 | \mathbf{y})}{P_{a_1}(o | \mathbf{y})} = \eta_{a_1}(\{2\} | \mathbf{y}) + \eta_{a_1}(\{1, 2\} | \mathbf{y}) = \eta_{a_1}(\{2\} | \mathbf{y}) + \eta_{a_1}(\{1\} | \mathbf{y})\eta_{a_1}(\{2\} | \mathbf{y}).$$

Hence, we identify  $\eta_{a_1}(\{2\} | \mathbf{y})$ . Repeating the above steps recursively we can identify  $\eta_{a_1}$  since

$$\eta_{a_1}(\{k\} | \mathbf{y}) = \frac{P_{a_1}(k | \mathbf{y})}{P_{a_1}(o | \mathbf{y})} \cdot \frac{1}{\sum_{\mathcal{C} \subseteq \{1, \dots, k-1\}} \eta_{a_1}(\mathcal{C} | \mathbf{y})}.$$

The same argument can be applied if  $\eta_{a_1}$  is additive. This approach can be generalized if one models the attention index of a set as a strictly increasing transformation of attention indexes of elements of that set (i.e.,  $\eta_{a_1}(\{1, 2\} | \mathbf{y}) = \phi_{\{1, 2\}}(\eta_{a_1}(\{1\} | \mathbf{y}), \eta_{a_1}(\{2\} | \mathbf{y}))$ , where  $\phi_{1,2}$  is strictly increasing in both arguments).

Suppose that  $\eta_{a_1}$  is the same for sets of the same cardinality. Since we know  $\eta_{a_1}(\{1\})$  from the first equation we identify attention indexes for all singleton sets including  $\eta_{a_1}(\{2\} | \mathbf{y})$ . Hence, using the second equation we identify  $\eta_{a_1}(\{1, 2\} | \mathbf{y})$  and, thus, attention indexes for all sets of cardinality 2. Repeating the above arguments recursively we can identify  $\eta_{a_1}$ .

It is left to show that if  $\eta_{a_1}$  is the same for sets that have the same best option, then it is also identified. Again,  $\eta_{a_1}(\{1\})$  is identified from the first equation. Since  $\eta_{a_1}(\{2\} | \mathbf{y}) = \eta_{a_1}(\{1, 2\} | \mathbf{y})$ , from the second equation we identify

$$\eta_{a_1}(\{2\} | \mathbf{y}) = \eta_{a_1}(\{1, 2\} | \mathbf{y}) = \frac{P_{a_1}(2 | \mathbf{y})}{2P_{a_1}(o | \mathbf{y})}.$$

Repeating the above arguments for every equation we get that

$$\eta_{a_1}(\mathcal{C} | \mathbf{y}) = \frac{P_{a_1}(v_{a_1, \mathcal{C}}^* | \mathbf{y})}{v_{a_1, \mathcal{C}}^* P_{a_1}(o | \mathbf{y})},$$

where  $v_{a_1, \mathcal{C}}^*$  is the best alternative in  $\mathcal{C}$  according to  $\succ_{a_1}$ . Since the choice of  $a_1$  and  $\mathbf{y}$  was arbitrary, we identify  $\eta_a$  for all  $a \in \mathcal{A}$ .

We conclude the proof by showing that restrictions (i) and (ii) correspond to the models of consideration sets formation in [Manzini and Mariotti \(2014\)](#) and [Dardanoni et al. \(2020\)](#), respectively. Equivalence between (ii) and the model in [Dardanoni et al. \(2020\)](#) is straightforward since

$$\frac{\eta_a(\mathcal{C} | \mathbf{y})}{\sum_{\mathcal{B} \subseteq \mathcal{Y}} \eta_a(\mathcal{B} | \mathbf{y})} = \frac{\eta_a(\mathcal{D} | \mathbf{y})}{\sum_{\mathcal{B} \subseteq \mathcal{Y}} \eta_a(\mathcal{B} | \mathbf{y})} \iff \eta_a(\mathcal{C} | \mathbf{y}) = \eta_a(\mathcal{D} | \mathbf{y}).$$

To show equivalence between (i) and the model in [Manzini and Mariotti \(2014\)](#) assume first that  $\eta_a(\cdot | \mathbf{y})$  is multiplicative (i.e., satisfies condition (i)). Then for every  $v \in \mathcal{Y}$  define

$$Q_a(v | \mathbf{y}) = \frac{\eta_a(\{y\} | \mathbf{y})}{1 + \eta_a(\{v\} | \mathbf{y})}.$$

Since  $\eta_a(\{v\} | \mathbf{y}) > 0$  (otherwise multiplicativity of  $\eta_a$  would imply that sets that contain  $v$  are never considered, which would contradict Assumption A1') and finite, we get that  $Q_a(v | \mathbf{y}) \in (0, 1)$ . Take any  $A$ . By multiplicativity of  $\eta_a$  we get that

$$\eta_a(A | \mathbf{y}) = \prod_{y \in A} \eta_a(\{y\} | \mathbf{y}).$$

Note that for singleton  $\mathcal{Y}$  we have that  $\sum_{C \subseteq \mathcal{Y}} \eta_a(C | \mathbf{y}) = 1 + \eta_a(\mathcal{Y} | \mathbf{y})$ . Then using multiplicativity we get that for  $\mathcal{Y} \cup \{y^*\}$  such that  $y^* \notin \mathcal{Y}$

$$\sum_{C \subseteq \mathcal{Y} \cup \{y^*\}} \eta_a(C | \mathbf{y}) = \sum_{C \subseteq \mathcal{Y}} \eta_a(C | \mathbf{y}) + \sum_{C \subseteq \mathcal{Y}} \eta_a(C \cup \{y^*\} | \mathbf{y}) = (1 + \eta_a(\{y^*\} | \mathbf{y})) \sum_{C \subseteq \mathcal{Y}} \eta_a(C | \mathbf{y}).$$

Hence, by induction

$$\sum_{C \subseteq \mathcal{Y}} \eta_a(C | \mathbf{y}) = \prod_{y \in \mathcal{Y}} (1 + \eta_a(\{y\} | \mathbf{y})).$$

As a result,

$$\begin{aligned} \frac{\eta_a(A | \mathbf{y})}{\sum_{C \subseteq \mathcal{Y}} \eta_a(C | \mathbf{y})} &= \prod_{y \in A} \frac{\eta_a(\{y\} | \mathbf{y})}{1 + \eta_a(\{y\} | \mathbf{y})} \prod_{y' \in \mathcal{Y} \setminus A} \left(1 - \frac{\eta_a(\{y'\} | \mathbf{y})}{1 + \eta_a(\{y'\} | \mathbf{y})}\right) \\ &= \prod_{y \in A} Q_a(y | \mathbf{y}) \prod_{y' \in \mathcal{Y} \setminus A} (1 - Q_a(y' | \mathbf{y})). \end{aligned}$$

Thus, condition (i) implies the model in [Manzini and Mariotti \(2014\)](#).

To show the opposite assume that the probability of considering a set  $\mathcal{A}$  is

$$\prod_{y \in \mathcal{A}} Q_a(y | \mathbf{y}) \prod_{y' \in \mathcal{Y} \setminus \mathcal{A}} (1 - Q_a(y' | \mathbf{y}))$$

for some  $Q_a$ . For singleton sets define  $\eta_a(\{y\} | \mathbf{y}) = \frac{Q_a(y | \mathbf{y})}{1 - Q_a(y | \mathbf{y})}$ . For sets with cardinality bigger than one define  $\eta_a(A | \mathbf{y}) = \prod_{y \in A} \eta_a(\{y\} | \mathbf{y})$ . By construction, the constructed  $\eta_a$  satisfies multiplicativity and it is easy to verify that

$$\prod_{y \in A} Q_a(y | \mathbf{y}) \prod_{y' \in \mathcal{Y} \setminus A} (1 - Q_a(y' | \mathbf{y})) = \frac{\eta_a(A | \mathbf{y})}{\sum_{C \subseteq \mathcal{Y}} \eta_a(C | \mathbf{y})}.$$

## B. Gibbs Random Field Model

In this section we connect our equilibrium concept with the so-called Gibbs equilibrium. This solution concept has been extensively used in Economics to study social interactions. (See [Allen, 1982](#), [Blume, 1993, 1995](#), and [Blume and Durlauf, 2003](#), among many others.) The starting point of the Gibbs random field models is a set of conditional probability distributions. Each element of the set describes the probability that a given person selects each alternative as a function of the profile of choices of the other people. In our model, the set of conditional probabilities is  $(P_a)_{a \in \mathcal{A}}$ , with a generic element given by

$$P_a(v | \mathbf{y}) = Q_a(v | y_a, N_a^v(\mathbf{y})) \prod_{v' \in \mathcal{Y}, v' \succ_a v} (1 - Q_a(v' | y_a, N_a^{v'}(\mathbf{y}))) \text{ for } v \in \mathcal{Y} \text{ and } \mathbf{y} \in \bar{\mathcal{Y}}^A.$$

A Gibbs equilibrium is defined as a joint distribution over the vector of choices  $\mathbf{y}$ ,  $P(\mathbf{y})$ , that is able to generate  $(P_a)_{a \in \mathcal{A}}$  as its conditional distribution functions.

Gibbs equilibria typically do not exist. (In the statistical literature, a similar existence problem is referred as the issue of compatibility of conditional distributions.) The existence of Gibbs equilibria depends on a great deal of homogeneity among people. In our model, it would also require  $Q_a(v | y_a, N_a^v(\mathbf{y}))$  to be invariant with respect to  $y_a$ . Condition G1 captures this restriction.

**(G1)** For each  $a \in \mathcal{A}$ ,  $v \in \mathcal{Y}$ , and  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ ,  $Q_a(v | \mathbf{y}) \equiv Q_a(v | N_a^v(\mathbf{y}))$ .

Together with Assumption A1, this extra condition allows a simple characterization of the invariant distribution  $\mu$ . We describe this characterization next.

**Proposition B.1.** *If A1 and G1 are satisfied, then there exists a unique  $\mu$ . Also,  $\mu$  satisfies*

$$\mu(\mathbf{y}) = \frac{1}{\sum_{a \in \mathcal{A}} \lambda_a} \sum_{a \in \mathcal{A}} \lambda_a P_a(y_a | \mathbf{y}) \mu_{-a}(\mathbf{y}_{-a}) \text{ for each } \mathbf{y} \in \bar{\mathcal{Y}}^A.$$

**Proof of Proposition B.1:** The characterization of  $\mu$  follows as the invariant distribution satisfies the balance condition  $\sum_{\mathbf{y}' \in \bar{\mathcal{Y}}^A} \mu(\mathbf{y}') m(\mathbf{y} | \mathbf{y}') = 0$  for each  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ . The next steps show this claim.

$$\sum_{\mathbf{y}' \in \bar{\mathcal{Y}}^A} \mu(\mathbf{y}') m(\mathbf{y} | \mathbf{y}') = 0$$

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

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

In what follows, we will normalize the lambdas to the value of 1. From Proposition B.1,  $\mu$  satisfies

$$\mu(\mathbf{y}) = \frac{1}{A} \sum_{a \in \mathcal{A}} P_a(y_a | \mathbf{y}) \mu_{-a}(\mathbf{y}_{-a}) \text{ for each } \mathbf{y} \in \bar{\mathcal{Y}}^A. \quad (5)$$

We next show that if  $(P_a)_{a \in \mathcal{A}}$  is a set of compatible conditional distributions, then  $\mu = P$  solves (5). If we let  $\mu = P$ , then right hand side of (5) is

$$\frac{1}{A} \sum_{a \in \mathcal{A}} P_a(y_a | \mathbf{y}) \sum_{v \in \bar{\mathcal{Y}}} P(v, \mathbf{y}_{-a}) = \frac{1}{A} \sum_{a \in \mathcal{A}} P(\mathbf{y}) = \frac{A}{A} P(\mathbf{y}) = P(\mathbf{y}).$$

Also, the left hand side of (5) is

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

Thus  $\mu(\mathbf{y}) = P(\mathbf{y})$  solves (5) for each  $\mathbf{y} \in \bar{\mathcal{Y}}^A$ .

As we mentioned earlier, and assumed in the last few lines, the existence of Gibbs equilibrium also requires the set of conditional probabilities  $(P_a)_{a \in \mathcal{A}}$  to be compatible. We formalize this idea next.

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

$$P_a(y_a | \mathbf{y}) = P(\mathbf{y}) / \sum_{y_a \in \bar{\mathcal{Y}}} P(\mathbf{y}) \text{ for each } \mathbf{y} \in \bar{\mathcal{Y}}^A.$$

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

$$\frac{1 - Q_1(1 | 0)}{Q_1(1 | 0)} \frac{Q_1(1 | 1)}{1 - Q_1(1 | 1)} = \frac{1 - Q_2(1 | 0)}{Q_2(1 | 0)} \frac{Q_2(1 | 1)}{1 - Q_2(1 | 1)}.$$

The last result states that, under specific conditions, the Gibbs equilibrium coincides with  $\mu$ . A similar connection is discussed in Blume and Durlauf (2003). The next proposition puts together



all our previous ideas.

**Proposition B.2.** *Assume A1 and G1 hold. If  $(P_a)_{a \in A}$  is a set of compatible conditional distributions, then  $P_a(y_a | \mathbf{y}) = \mu(\mathbf{y}) / \mu_{-a}(\mathbf{y}_{-a})$  for each  $\mathbf{y} \in \mathcal{Y}^A$ .*

We next use our first example to illustrate this result.

**Example 1 (continued).** The conditional choice probabilities in Example 1 satisfy the compatibility requirements. Thus, there exists a joint distribution on  $y_1$  and  $y_2$  that can generate them as its conditional distribution functions. In this simple case, the invariant distribution  $\mu$  of the dynamic revision process coincides with the Gibbs equilibrium of the model. To see this notice that from our previous results we get

$$\begin{aligned} \mu(o) &= \mu(o, o) + \mu(o, 1) = \mu(o, o) + \mu(1, o) = \frac{1 - Q(1 | 1)}{1 - Q(1 | 1) + Q(1 | 0)}, \\ \mu(1) &= \mu(1, 1) + \mu(o, 1) = \mu(1, 1) + \mu(1, o) = \frac{Q(1 | 0)}{1 - Q(1 | 1) + Q(1 | 0)}. \end{aligned}$$

(Here again, we write  $Q$  instead of  $Q_a$  due to the symmetry across people.) Thus, the conditional distributions are

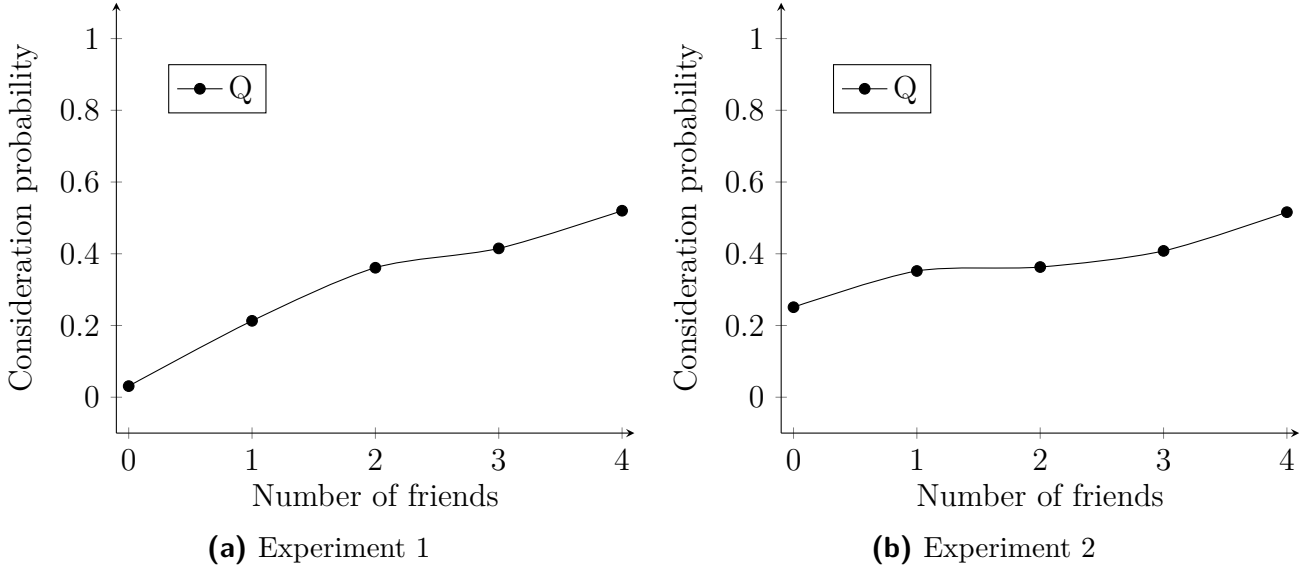
$$\begin{aligned} \mu(1, o) / \mu(o) &= Q(1 | 0) \quad \text{and} \quad \mu(o, o) / \mu(o) = 1 - Q(1 | 0) \\ \mu(1, 1) / \mu(1) &= Q(1 | 1) \quad \text{and} \quad \mu(o, 1) / \mu(1) = 1 - Q(1 | 1). \end{aligned}$$

It follows that the conditional choice probabilities generated by equilibrium behavior coincide with the conditional choice probabilities in the model. ■

## C. Omitted Details of Monte Carlo Simulations

This appendix describes how we generated the data for our simulations. 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  $\lambda_a/\lambda$ ;
- (iv) Given the agent selected in the previous step and the current choice configuration  $\mathbf{y}_k$  construct a consideration set using  $Q_a$ ;



**Figure 3** – Consideration probability. Model Ia. Middle half of the samples.

- (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  $y_{a,k}$ .

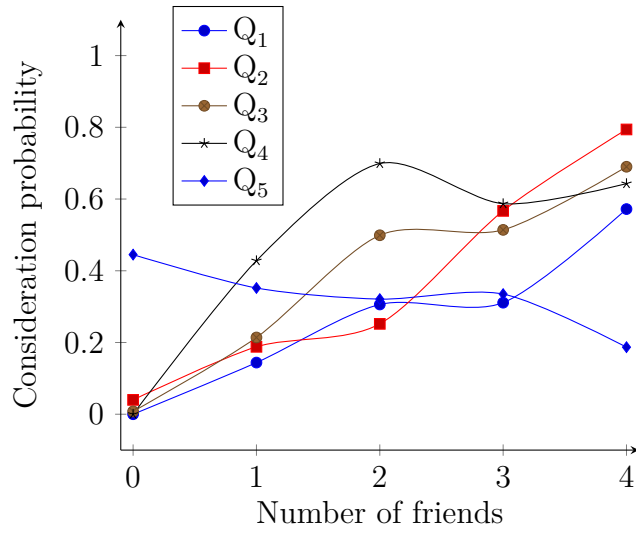
Given the initial configuration of choices  $\mathbf{y}_0$  we applied the above algorithm till we reached  $\sum_k x_k > \mathcal{T}$  (On average the length of the sequence is  $\lambda\mathcal{T}$ ). Define  $z_k = \sum_{l \leq k} x_l$ . The continuous time data is  $\{(y_k, z_k)\}$ . The discrete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

## D. Additional Estimation Results

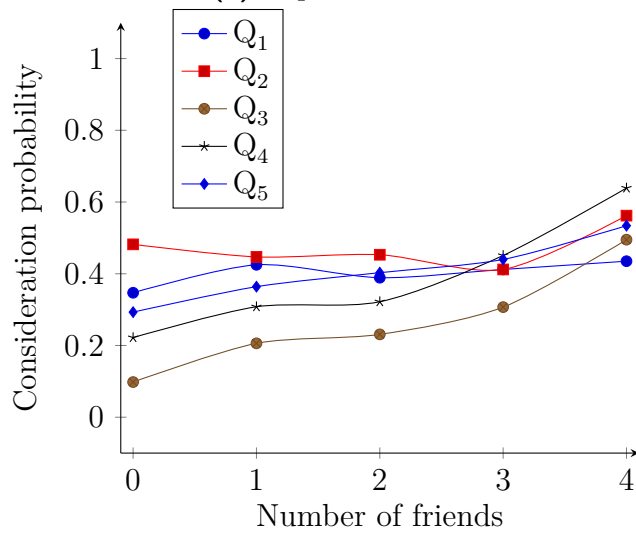
### D.1. Different Parts of the Sample

Given that the Resistance is a dynamic party game, in order to analyse the behavior of consideration probabilities and preferences over time we estimated Models I(a) and I(b) presented in Section 7 using two different subsamples. The first subsample consists of the observations coming from the middle half (i.e., we threw away the first and the last quarter of observations). The second one contains observations from the second half.

The estimated consideration set probabilities for the first subsample are presented in Figures 3-6. The estimated preference orders are presented in Tables 9-10.

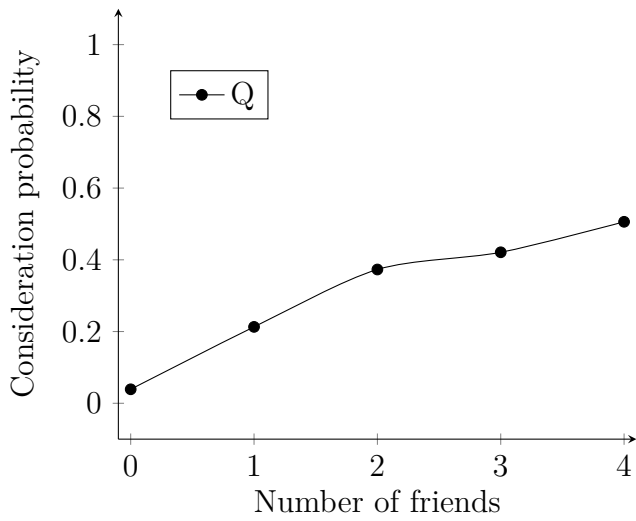


(a) Experiment 1

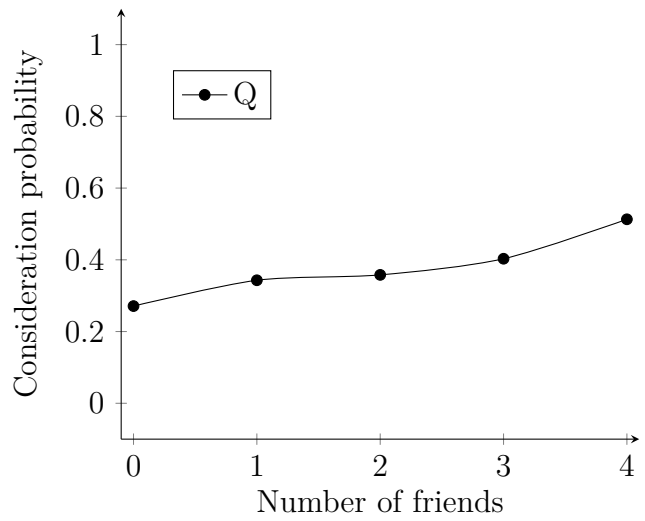


(b) Experiment 2

Figure 4 – Consideration probabilities. Model I(b). Middle half of the samples.

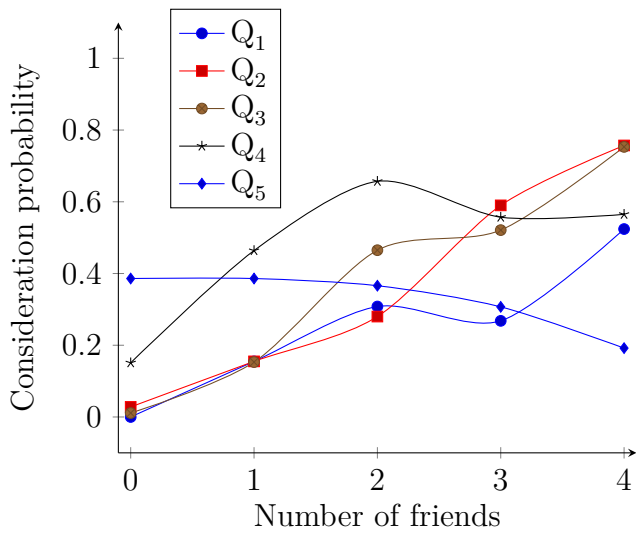


(a) Experiment 1

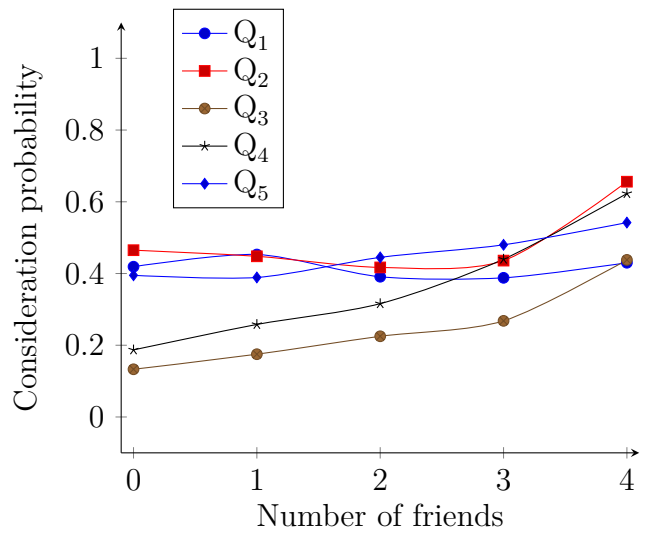


(b) Experiment 2

Figure 5 – Consideration probability. Model Ia. Second half of the samples.



(a) Experiment 1



(b) Experiment 2

Figure 6 – Consideration probabilities. Model I(b). Second half of the samples.

**Table 9** – Preferences. Model I(a).

		Player 1	Player 2	Player 3	Player 4	Player 5
Middle half	Experiment 1	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R
	Experiment 2	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R
2nd half	Experiment 1	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R
	Experiment 2	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R

**Table 10** – Preferences. Model I(b).

		Player 1	Player 2	Player 3	Player 4	Player 5
Middle half	Experiment 1	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R
	Experiment 2	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R
2nd half	Experiment 1	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R
	Experiment 2	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R

## D.2. Heterogeneity in Consideration Probabilities Across Choices

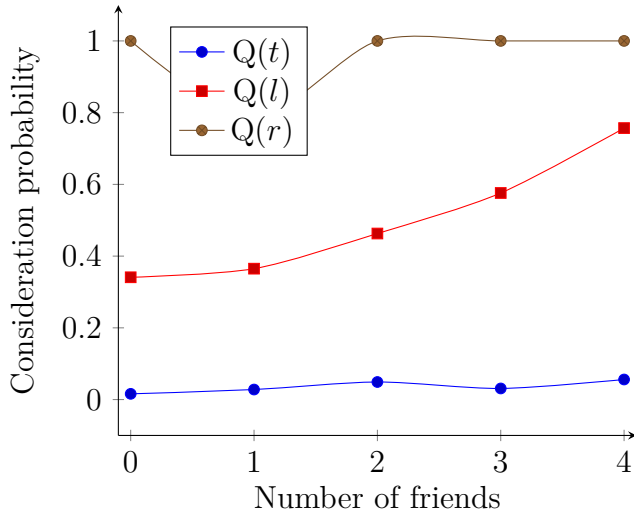
In this section we present results of estimation of the following specifications of the model:

- (i) Model Ic:  $Q_a(v | y_a, N_a^v(\mathbf{y})) = Q(v | N_a^v(\mathbf{y}))$  for all  $a, v, \mathbf{y}$ ;
- (ii) Model IIa:  $Q_a(v | y_a, N_a^v(\mathbf{y})) = Q(\text{State}, N_a^v(\mathbf{y}))$  for all  $a, v, \mathbf{y}$ . State equals to 1 if  $v = y_a$ , and 0 otherwise.
- (iii) Model IIb:  $Q_a(v | y_a, N_a^v(\mathbf{y})) = Q_a(\text{State}, N_a^v(\mathbf{y}))$  for all  $a, v, \mathbf{y}$ . State equals to 1 if  $v = y_a$ , and 0 otherwise.

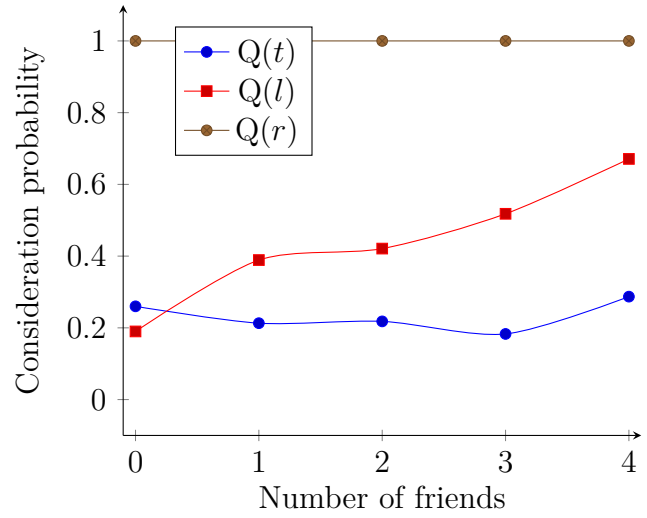
The estimated consideration probabilities are presented in Figures 7- 10 and Tables 11- 13.

**Table 11** – Preferences. Model Ic

	Player 1	Player 2	Player 3	Player 4	Player 5
Experiment 1	T,L,R	T,L,R	L,T,R	T,L,R	L,T,R
Experiment 2	T,L,R	T,L,R	L,T,R	T,L,R	T,L,R

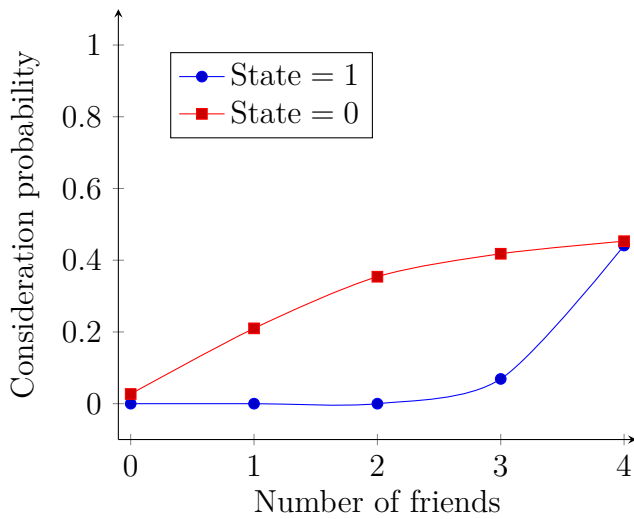


(a) Experiment 1

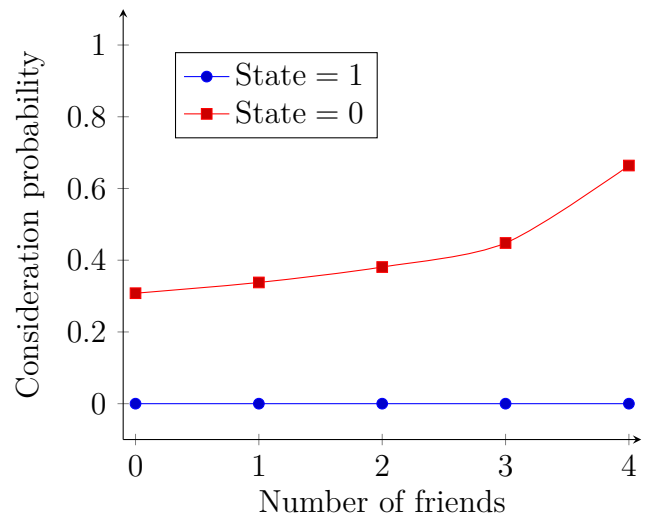


(b) Experiment 2

Figure 7 – Consideration probabilities. Model Ic.

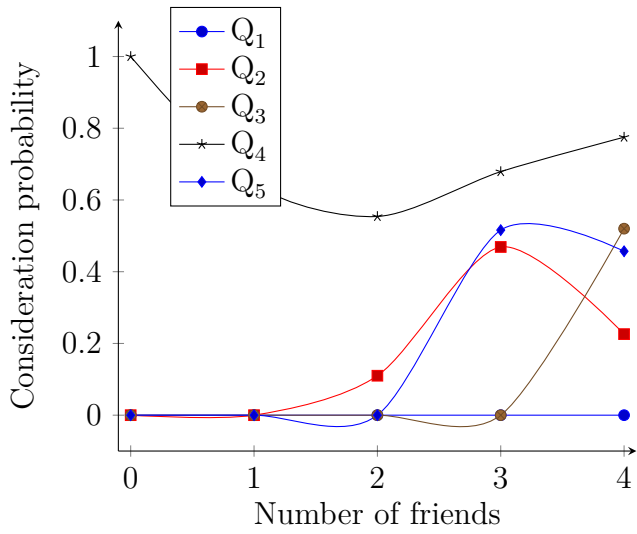


(a) Experiment 1

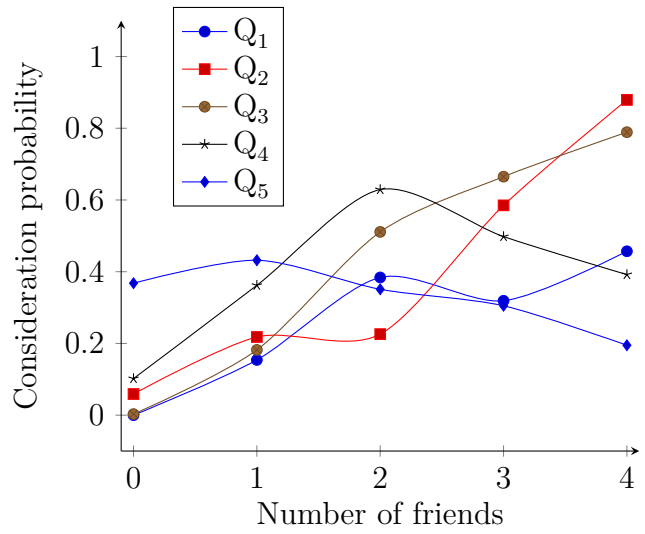


(b) Experiment 2

Figure 8 – Consideration probabilities. Model IIa.

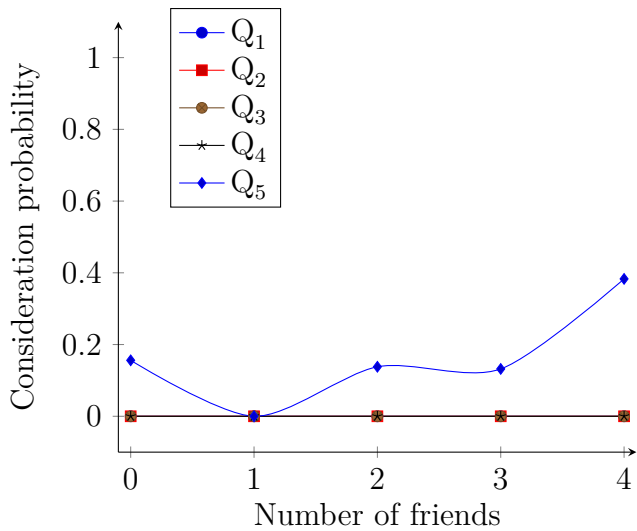


(a) State=1

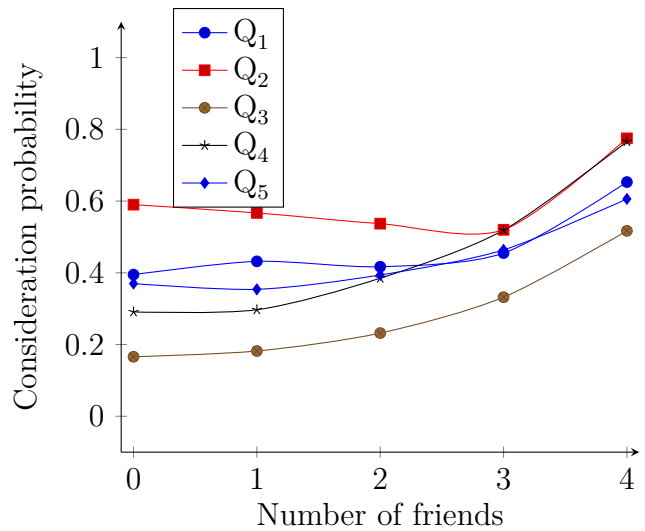


(b) State=0

**Figure 9** – Consideration probabilities. Model IIb. Experiment 1.



(a) State=1



(b) State=0

**Figure 10** – Consideration probabilities. Model IIb. Experiment 2.

**Table 12** – Preferences. Model IIa

	Player 1	Player 2	Player 3	Player 4	Player 5
Experiment 1	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R
Experiment 2	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R

**Table 13** – Preferences. Model IIb

	Player 1	Player 2	Player 3	Player 4	Player 5
Experiment 1	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R
Experiment 2	L,T,R	L,T,R	L,T,R	L,T,R	L,T,R