Peer Effects in Random Consideration Sets*

Nail Kashaev — Western University — nkashaev@uwo.ca

Natalia Lazzati — UC Santa Cruz — nlazzati@ucsc.edu

This version: September 12, 2022

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. We apply our results to an experimental dataset that has been designed to study the visual focus of attention. We find robust left-to-right bias and positive peer effects in gazing.

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, Seteven Durlauf, Paola Manzini, Marco Mariotti, Salvador Navarro, and Al Slivinski. 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, the Econometrics of Games, Matching and Network at Toulouse, Bristol ESG, and the seminar presentations at Cornell University, UC Riverside, and 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 people's choices. 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.[1] To study this possibility, we build a dynamic model in which the choices of friends affect the subset of options that a person ends up considering. We show that all parts of the model can be recovered from a sequence of choices, and apply our results to an experimental dataset that studies the visual focus of attention of people. The results corroborate existing beliefs on the left-to-right bias in visual focus of attention and peer effects in gazing. They also highlight the relevance of a structural approach to recovering preferences.

In our model, 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. The person does not consider all the available options when revising her selection. Instead, she first forms a consideration set and then picks her most preferred option from it. The distinctive feature of our model is that the probability that a given alternative enters the consideration set depends on the number of friends currently adopting that option. This model leads to a sequence of choices that evolves through time according to a Markov random process. We show that this dynamic process has a unique equilibrium and describe the equilibrium behavior.

The model we build might fit in various applications. As a first one, let us consider an online platform that offers many video games to a set of users.[2] (These games could include Super Mario

[1] 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, 2014), a new retirement plan (Duflo and Saez, 2003), a new restaurant (Qiu, Shi and Whinston, 2018), or an opportunity to protest (Enikolopov, Makarin and Petrova, 2020).

[2] See, Lee (2015) for an example.

Bros, Castlevania, or Sonic the Hedgehog.) A user can enter the platform and decide which game to get, at a given price, if any. The number of games offered by the platform is often quite large. The platform could create a reference group for each user and share with her the last purchasing decisions of the other group members. In this situation, it might be natural to think the information will affect the subset of games to consider. Our model can help the platform personalize each user's reference group to maximize profits or the probability of making a sale.

As a second application, note that we can re-think our model as a model of peer effects in mistakes. To illustrate this possibility, let us say that a person makes a mistake when, at the moment of choosing, she does not consider, and thereby does not select, her most preferred alternative. In our setting, the frequency by which a person makes mistakes depends on her friends' choices. We show via example that having friends with similar preferences – a phenomenon known as homophily – often reduces the frequency by which people make mistakes. We also offer conditions under which having more friends benefits the person.

After we show equilibrium existence and characterize equilibrium behavior, we consider a researcher who observes a long sequence of choices made by the network members. 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. 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 probability 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 and the variation in consideration set probabilities induced by choices of friends. Second, we study identification of the conditional choice probabilities from observed data.

The identification strategy we propose for the primitives of the model is simple and constructive. To recover them 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 currently choosing it. As we just mentioned, this identification result 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 vector of choices at fixed time intervals (e.g., the choices of people are observed every Monday). In this case, the conditional choice probabilities are not directly observed, 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). The second dataset suffices to recover the conditional choices because 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 simultaneously is zero. Thus, the transition rate matrix has zeros in many known locations. The non-zero elements can then be 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; we let the distribution of consideration sets be multiplicatively separable across alternatives; and we let the probability of including each option depend on the number (but not the identity) of the friends that selected that option. We then 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 option. All the primitives of this model are identified if there are more than three options in the set of available alternatives. Third, we allow the default option not to be fully dominated by all other alternatives, and its ranking to differ across people. This extension is particularly important for applications such as the online platform we described earlier. 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.

Last, but not least, we apply our model to an experimental dataset that has been used by Bai, Kumar, Leskovec, Metzger, Nunamaker Jr and Subrahmanian (2019) to compare the effectiveness of various models that aim to predict the visual focus of attention of individuals within a group. In the experiment, a group of people were asked to play a party game in which they communicated with each other to find the deceptive players. The players were seated around a circle, and a tablet, which was placed in front of each player, recorded the direction of the player's sight every third of a second. We consider each person to be looking to the left, to the tablet, or to the right. The raw frequencies (in percent) of the direction of sight of the five players in the data are as follows.

	Player 1	Player 2	Player 3	Player 4	Player 5
left	40	35	30	29	50
tablet	20	38	5	15	16
right	40	27	65	56	34

The approach we offer in the paper allows us to separate two key forces in explaining the data:

directional sight preferences of players and peer effects in gazing. These effects 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, everything else equal, people spend more time looking to the left side of the screen as compared to the right side. This phenomenon is called the left-to-right bias. 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. Although, the choice set is not large, the participants had to make decisions in seconds, thus, justifying limited consideration. The empirical estimates we obtain are striking. In particular, despite considerable differences in the raw frequency of the three choices across players, as reflected in the above table, the estimated preferences for directional sight of all players coincide and are consistent with the left-to-right bias. Also, the estimates for the peer effects in consideration sets (in the data) are positive and relevant. These findings highlight the importance of incorporating peers' inattention behavior and individual preferences as different channels for decision making.

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 the variation of the set of available options or menus. The latter includes Aguiar (2017), Aguiar, Boccardi and Dean (2016), Brady and Rehbeck (2016), Caplin, Dean and Leahy (2019), Cattaneo, Ma, Masatlioglu and Suleymanov (2020), Horan (2019), Kashaev ⓡ Aguiar (2022), Lleras, Masatlioglu, Nakajima and Ozbay (2017), Manzini and Mariotti (2014), and Masatlioglu, Nakajima and Ozbay (2012). (See Aguiar, Boccardi, Kashaev and Kim, 2022 for a comparison of several consideration set models in an experiment.) Other papers have relied on exogenous covariates that shift preferences or consideration sets. The latter include Barseghyan, Molinari and Thirkettle (2021b), Barseghyan, Coughlin, Molinari and Teitelbaum (2021a), Crawford, Griffith and Iaria (2021), Conlon and Mortimer (2013), Draganska and Klapper (2011), Gaynor, Propper and Seiler (2016), Goeree (2008), Mehta, Rajiv and Srinivasan (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, Manzini, Mariotti and Tyson (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, Brock, Durlauf and Ioannides, 2011, Bramoullé, Djebbari and Fortin, 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.[3] Moreover, our model leads to different empirical predictions (see Appendix C for further details). We view our work as complementing the existing results on peer effects in preferences by proposing an alternative mechanism for the social interaction effects. 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, Brock, Durlauf and Jayaraman (2015), Bonaldi, Hortaçsu and Kastl (2015), De Paula, Rasul and Souza (2018), and Manresa (2013). In the context of discrete-choice, Chambers, Cuhadaroglu and Masatlioglu (2019) also identifies the network structure. However, 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 from 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

[3]Implicitly, preferences in our model capture the so-called contextual effects – effects of predetermined factors that are not affected by peers.

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. 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), two variants of the default option, and more general formation processes for the consideration sets. Section 6 applies our model to an experimental dataset on visual focus of attention. Section 7 concludes, and all the proofs are collected in Appendix A. Appendix B provides some simulation evidence of finite sample performance of our estimator. In Appendices C and D we compare our framework to a model of peer effects in preferences and the Gibbs random field model, respectively. Appendix E contains additional results for our empirical application.

2. The Model

This section describes the model and the main assumptions we invoke in the paper. It also elaborates on the notion of mistakes we will use later.

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 $\mathcal{A} = \{1, 2, \ldots, A\}$ 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 $\overline{\mathcal{Y}} = \mathcal{Y} \cup \{o\}$, where $\mathcal{Y} = \{1, 2, \ldots, Y\}$ 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. These assumptions are standard in the literature on single-agent choice with limited consideration and allow us connect our framework with that literature. We extend the analysis to random preferences in Section 5.1, relax the specification of the default option in Section 5.2, and consider the case in which people differ regarding the ranking in preferences of the default option in Section 5.3. We refer to a vector $\mathbf{y} = (y_a)_{a \in \mathcal{A}} \in \overline{\mathcal{Y}}^A$ 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 $\lambda = (\lambda_a)_{a \in \mathcal{A}}$.[4] At randomly given moments (exponentially distributed with mean $1/\lambda_a$) the alarm of Person a goes off.[5] 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

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

that takes value 1 if v is the most preferred option in $\mathcal{C}$ according to $\succ_a$ and is 0 otherwise. 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 of her friends at the moment of revising her selection. We indicate by $\mathrm{Q}_a (v \mid \mathbf{y})$ the probability that Person a pays attention to alternative v given a choice configuration $\mathbf{y}$. It follows that the

[4]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.

[5]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.

probability of facing consideration set $\mathcal{C}$ has the form of

$$\prod_{v\in\mathcal{C}} \mathrm{Q}_a(v \mid \mathbf{y}) \prod_{v\notin\mathcal{C}} (1 - \mathrm{Q}_a(v \mid \mathbf{y})).$$

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

$$\mathrm{P}_a(v \mid \mathbf{y}) = \mathrm{Q}_a(v \mid \mathbf{y}) \prod_{v'\in\mathcal{Y}, v' \succ_a v} (1 - \mathrm{Q}_a(v' \mid \mathbf{y})). \tag{1}$$

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 $\prod_{v\in\mathcal{Y}} (1 - \mathrm{Q}_a(v \mid \mathbf{y}))$. 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 5.4 we extend our analysis to a more general setting. Note that, by construction, the peer effects in our model do not involve forward-looking behavior. This modeling restriction is not introduced for technical reasons. We do so because we want to capture situations where the choices of friends simply make a person more aware of the existence of specific alternatives or redirect her attention in the direction of that option.

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. We next present three key assumptions we invoke in the paper.

2.2. Main Assumptions

Our results build on one or more of the three assumptions we discuss next. Let $\mathrm{N}_a^v(\mathbf{y})$ be the number of friends of Person a who select option v in choice configuration $\mathbf{y}$. Formally,

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

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 $\mathbf{y} \in \overline{\mathcal{Y}}^A$, $1 > \mathrm{Q}_a(v \mid \mathbf{y}) > 0$.

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

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

$$\mathrm{Q}_a(v \mid \mathbf{y}) \equiv \mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})) \text{ is (weakly) increasing in } \mathrm{N}_a^v(\mathbf{y})$$
$$\text{with strict monotonicity at } \mathrm{N}_a^v(\mathbf{y}) = 0.$$

Assumption A1 states that 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). In line with the spirit of the limited attention mechanisms, 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. The monotonicity assumption is required to be strict only at zero, allowing for different levels of satiation —e.g. consideration changes only when the number of friends picking the option achieves a given threshold (e.g., 10 friends, 20 friends, etc).[6]

Note that Assumptions A1 and A3 allow a setting where the current selection is considered with probability close to 1, thus capturing environments with inertia. These modelling restrictions could, nevertheless, be further relaxed. For instance, we could incorporate dependence of the consideration sets not just on current choices, but also on past decisions —e.g., a Markov process with memory. This extension could be achieved by enlarging the state space. Let us also note that we assume

[6] Assumption A3 can be relaxed. The strict monotonicity does not need to hold at zero or be the same for all agents, as long as it is below $|\mathcal{N}_a|$. We write the assumption in the stricter way to simplify the exposition.

peer effects in consideration sets are positive as this assumption is consistent with the sort of applications that motivated our work. Nevertheless, it is interesting to remark that our results still hold if peer effects were negative for some people and positive for other persons —these results include both equilibrium existence and identification of the model. Moreover, we could be agnostic about the signs of the peer effects and recover them (separately) for each person from the data.

2.3. Applications: Peer Effects in Mistakes and Sales of an Online Platform

We think our model might be particularly appealing in specific situations. We describe two of them next.

Peer Effects in Mistakes In our model, Person a has a strict preference order over the set of available alternatives, and the choices of her friends affect the probability that the person considers each particular option. We say that Person a makes a mistake when she does not select her most preferred alternative at the moment of choosing. In our setting, this happens when the person does not include that alternative in her consideration set. More formally, let v_a^* be the most preferred alternative of Person a in $\mathcal{A}$ according to $\succ_a$. Given our definition, the probability that Person a makes a mistake at the moment of choosing an option is given by

$$\mathrm{Mistake}_a\left(\mathbf{y}\right) = 1 - \mathrm{Q}_a\left(v_a^* \mid y_a, \mathrm{N}_a^{v_a^*}\left(\mathbf{y}\right)\right).$$

This probability depends on the number of friends who are currently selecting her most preferred alternative. We thereby can think of our model as a model of peer effects in mistakes. The possibility of reinterpreting the model in this way makes it clear a key difference between the models of peer effects in consideration sets and the one in preferences. We will show later how the strength of people's mistakes depends on the structure of the network.

Sales of an Online Platform Let us now think that the different alternatives are offered by an online platform at some prices. Under this interpretation of the model, we can think of the default option as the possibility of not buying any alternative. The platform could be interested in minimizing this probability. Since the default is picked if and only if nothing else is considered, for each Person a, the probability of selecting the not-to-but-any-alternative is given by

$$\mathrm{P}_a(o \mid \mathbf{y}) = \prod\nolimits_{v \in \mathcal{Y}} \left(1 - \mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})\right)\right).$$

This probability depends on the number of friends that are selecting each of the available options. The platform could affect this probability by proposing specific links to each person. Indeed, many platforms are currently implementing this practice. In Section 5.3, we extend the initial model to allow the default option to have a different ranking in the order of preferences of each person. This extension might be particularly useful in this type of application.

3. Equilibrium Behavior

This section states equilibrium existence and offers some insights on equilibrium behavior. (In Appendix C we show how our model differs from the more standard model of peer effects in preferences.)

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) offer 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

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

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

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

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

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

An equilibrium in this model is an invariant distribution $\mu : \overline{\mathcal{Y}}^{A} \to [0,1]$, with $\sum_{\mathbf{y} \in \overline{\mathcal{Y}}^{A}} \mu\left(\mathbf{y}\right) = 1$, 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 μ.*

Remark. In Appendix D we relate the equilibrium concept we use with the so-called Gibbs equilibrium, which 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 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., $\mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})\right) = \mathrm{Q}_a\left(v \mid \mathrm{N}_a^v(\mathbf{y})\right)$) and is the same for both people (i.e., $\mathrm{Q}(\cdot) = \mathrm{Q}_1(\cdot) = \mathrm{Q}_2(\cdot)$). Thus, for $a = 1, 2$, we get that

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

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

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

.

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.

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

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

The 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. ■

Recall we said that a person makes a mistake when she does not select, because she does not consider, her most preferred option at the moment of choosing. The frequency of mistakes that each person makes in the long run depends on the distribution of choices at equilibrium. Formally,

$$\text{EqMistake}_a = 1 - \mu_a\left(v_a^*\right),$$

where v_a^* is the most preferred alternative of Person a in $\mathcal{A}$ according to $\succ_a$. The second example shows how the frequencies of mistakes (which people make in equilibrium) are related to the structure of the network.

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 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 $\mathrm{Q}\left(v \mid \mathrm{N}_a^v\left(\mathbf{y}\right)\right)$ 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 $\mathrm{N}_a^v\left(\mathbf{y}\right)$. We let

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

The rates for their Poisson alarm clocks are 1.

The equilibrium behavior of this choice model is a joint distribution μ with support on 243 choice configurations (3^5). We calculated the equilibrium behavior. (See Appendix B for more details.) Table 1 displays the marginal frequency of choices of each person (at equilibrium).

Notice that for some people in the network, the relative frequency of choosing each alternative

Table 1 – Marginal Equilibrium Probabilities

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

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

Table 2 – Mistakes in Equilibrium

	Person 1	Person 2	Person 3	Person 4	Person 5
EqMistakes	59%	60%	49%	50%	50%

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

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

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. Specifically, Person 4 shares with her friend the same preference order. The opposite is true for Person 2. In this application, this difference leads Person 4 to make fewer mistakes. Hence, having friends with similar preferences (homophily) seems to help people to consider more often their best alternatives and to make fewer mistakes.

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 in the consideration sets 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 and the probabilities of mistakes depicted in Tables 3 and 4, respectively.

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. ■

In Example 2, Persons 1 and 2 reduce the frequency of mistakes when Person 3 is added to their

Table 3 – Marginal Equilibrium Probabilities

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

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

Table 4 – Mistakes in Equilibrium

	Person 1	Person 2	Person 3	Person 4	Person 5
EqMistakes	33%	58%	34%	50%	50%

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

reference groups. This result might suggest that having more friends is always helpful in reducing the frequency of mistakes. The next proposition offers precise conditions under which this is the case. We then show, via example, that some extra conditions are needed for this result to hold. That is, this result might fail otherwise.

Proposition 3.2. *In addition to Assumptions A1 and A3 suppose that* $\mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})) = \mathrm{Q}_a(v \mid \mathrm{N}_a^v(\mathbf{y}))$ *for all* $v \in \mathcal{Y}$*, and* $a \notin \mathcal{N}_{a'}$ *for all* $a' \in \mathcal{A}$*. Let* $EqMistake_a$ *and* $EqMistake'_a$ *be the frequency of mistakes at equilibrium for Person* a *when her reference group is* $\mathcal{N}_a$ *and* $\mathcal{N}'_a$*, respectively. If* $\mathcal{N}'_a \subseteq \mathcal{N}_a$*, then we have that*

$$\mathrm{EqMistake}'_a \geq \mathrm{EqMistake}_a\,. \tag{3}$$

Proposition 3.2 requires that Person a does not enter the consideration set of others. In this case, when the reference group of Person a gets larger, the probability of paying attention to each alternative moves up. Since among these alternatives is her most preferred option, then the person makes fewer mistakes. The next example shows that, without extra restrictions, having more friends might lead to making more mistakes.

Example 3. There are three people in the network. Their reference groups are as follows

$$\mathcal{N}_1 = \{2\}\,, \quad \mathcal{N}_2 = \{1\}\,, \quad \mathcal{N}_3 = \{\emptyset\}\,.$$

Table 5 – Marginal Equilibrium Probabilities

$\mu_1(o) = 0.48$	$\mu_2(o) = 0.54$	$\mu_3(o) = 0.18$
$\mu_1(1) = 0.37$	$\mu_2(1) = 0.40$	$\mu_3(1) = 0.80$
$\mu_1(2) = 0.15$	$\mu_2(2) = 0.06$	$\mu_3(2) = 0.02$

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

Table 6 – Mistakes at Equilibrium

	Person 1	Person 2	Person 3
EqMistakes	85%	60%	20%

Notes: These probabilities are computed using the estimated marginal equilibrium probabilities from Table 5.

Persons 1 and 2 are interconnected and Person 3 is isolated. There are two possible alternatives $\mathcal{Y} = \{1, 2\}$ in addition to the default option o. The preferences of these people are as follows

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

That is, Person 1 prefers option 2 and Persons 2 and 3 prefer option 1. Similar to the previous examples, we assume the probability of paying attention to a given option only depends on the choices of friends. In this case, we indicate by $\mathrm{Q}_a\left(v \mid \mathrm{N}_a^v(\mathbf{y})\right)$ 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 $\mathrm{N}_a^v(\mathbf{y})$. We let

$$\begin{aligned}
&\mathrm{Q}_1(1 \mid 0) = \tfrac{1}{10},\ \mathrm{Q}_1(1 \mid 1) = \tfrac{9}{10},\ \mathrm{Q}_1(1 \mid 2) = \tfrac{9}{10},\ \mathrm{Q}_1(2 \mid 0) = \tfrac{1}{10},\ \mathrm{Q}_1(2 \mid 1) = \tfrac{8}{10},\ \mathrm{Q}_1(2 \mid 2) = \tfrac{9}{10}\\
&\mathrm{Q}_2(1 \mid 0) = \tfrac{1}{10},\ \mathrm{Q}_2(1 \mid 1) = \tfrac{9}{10},\ \mathrm{Q}_2(1 \mid 2) = \tfrac{9}{10},\ \mathrm{Q}_2(2 \mid 0) = \tfrac{1}{10},\ \mathrm{Q}_2(2 \mid 1) = \tfrac{1}{10},\ \mathrm{Q}_2(2 \mid 2) = \tfrac{1}{10}\\
&\mathrm{Q}_3(1 \mid 0) = \tfrac{8}{10},\ \mathrm{Q}_3(1 \mid 1) = \tfrac{9}{10},\ \mathrm{Q}_3(1 \mid 2) = \tfrac{9}{10},\ \mathrm{Q}_3(2 \mid 0) = \tfrac{1}{10},\ \mathrm{Q}_3(2 \mid 1) = \tfrac{1}{10},\ \mathrm{Q}_3(2 \mid 2) = \tfrac{1}{10}.
\end{aligned}$$

The rates for the Poisson alarm clocks are 1. Tables 5 and 6 display the marginal distributions of equilibrium choices and mistakes, respectively.

Table 7 – Marginal Equilibrium Probabilities

$\mu_1(o) = 0.13$	$\mu_2(o) = 0.28$	$\mu_3(o) = 0.18$
$\mu_1(1) = 0.73$	$\mu_2(1) = 0.68$	$\mu_3(1) = 0.80$
$\mu_1(2) = 0.14$	$\mu_2(2) = 0.04$	$\mu_3(2) = 0.02$

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

Table 8 – Mistakes at Equilibrium

	Person 1	Person 2	Person 3
EqMistakes	86%	31%	20%

Notes: These probabilities are computed using the estimated marginal equilibrium probabilities from Table 7.

Let us now add Person 3 to the consideration set of Person 1. That is,

$$\mathcal{N}_1 = \{2, 3\}, \quad \mathcal{N}_2 = \{1\}, \quad \mathcal{N}_3 = \{\emptyset\}.$$

Repeating the previous exercise, the new network generates the marginal equilibrium distributions and mistakes depicted in Tables 7 and 8, respectively.

Notice that while we only increased the reference group of Person 1, her frequency of mistakes at equilibrium moved up. The reason for this result is that after Person 3 was added to the consideration set of Person 1, Person 1 increased the frequency of selecting option 2. In turn, this lead Person 2 to consider this option more often. Thus, while Person 1 sharply decreased the frequency by which she selected the default option, she simultaneously increased the frequency of selecting option 2 and decreased the frequency of option 1. ■

We could easily construct similar examples for the sales application of the online platform. We prefer not to do so to shorten the exposition. But a lesson to learn from the last results is that if the platform wants to offer reference groups to each user to maximize sales, it has to select them quite carefully. The identification results in the next section could guide the firm to implement an optimal (or profit maximizing) selling strategy.

4. Identification

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 $\mathrm{Q} = (\mathrm{Q}_a)_{a\in\mathcal{A}}$, and the rates of the Poisson alarm clocks $\lambda = (\lambda_a)_{a\in\mathcal{A}}$. In Appendix B, we propose a maximum likelihood estimator of the model parameters and conduct several Monte Carlo experiments to evaluate its finite-sample properties.

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

$$\mathrm{P}_a(v \mid \mathbf{y}) = \mathrm{Q}_a(v \mid \mathbf{y}) \prod\nolimits_{v'\in\mathcal{Y}, v'\succ_a v} (1 - \mathrm{Q}_a(v' \mid \mathbf{y}))$$

and the probability of selecting the default option o is $\prod_{v\in\mathcal{Y}} (1 - \mathrm{Q}_a(v \mid \mathbf{y}))$. 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 show that the conditional choice probabilities P can be recovered 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, A2, and A3, the set of connections* Γ*, 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 4. Suppose that three people $\mathcal{A} = \{1, 2, 3\}$ 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

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

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

$$\mathrm{P}_1\left(o \mid o, o, o\right) > \mathrm{P}_1\left(o \mid o, 1, o\right).$$

In words, 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 $\mathrm{P}_1\left(o \mid o, o, o\right) > \mathrm{P}_1\left(o \mid o, o, 1\right)$. 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

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

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

$$\mathrm{P}_1\left(1 \mid o, o, o\right) > \mathrm{P}_1\left(1 \mid o, 2, o\right).$$

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

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

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. ■

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 Δ and can consistently estimate $\Pr\left(\mathbf{y}^{t+\Delta} = \mathbf{y}' \mid \mathbf{y}^t = \mathbf{y}\right)$ for each pair $\mathbf{y}', \mathbf{y} \in \overline{\mathcal{Y}}^A$. We will capture these transition probabilities by a matrix $\mathcal{P}(\Delta)$. (Here again, we will assume that the choice configurations are ordered according to the lexicographic order when we construct $\mathcal{P}(\Delta)$.) The connection between $\mathcal{P}(\Delta)$ and 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 Δ: 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 $\lim_{\Delta \to 0} \mathcal{P}(\Delta)$
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* P *and the rates of the Poisson alarm clocks* λ *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 $\Delta \to 0$. One of them requires the length of the interval Δ to be below a threshold $\overline{\Delta}$. 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 Δ_1 and Δ_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* λ*, 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* 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 $\mathrm{R}_a(\cdot \mid \mathcal{C})$ from Section 2 is not an indicator function but a distribution on $\mathcal{Y}$. We naturally let $\mathrm{R}_a(v \mid \mathcal{C}) = 0$ 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

$$\mathrm{P}_a(v \mid \mathbf{y}) = \sum_{\mathcal{C} \subseteq 2^{\mathcal{Y}}} \mathrm{R}_a(v \mid \mathcal{C}) \prod_{v' \in \mathcal{C}} \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right) \prod_{v' \notin \mathcal{C}} \left(1 - \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right)\right). \quad (4)$$

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

$$\mathrm{R}_a(1 \mid \{1, 2\}) = \frac{\exp(\mathrm{U}_a^1)}{\exp(\mathrm{U}_a^1) + \exp(\mathrm{U}_a^2)}.$$

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* Γ *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 $\mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v\left(\mathbf{y}\right)\right) / \mathrm{Q}_a\left(v \mid y_a, 0\right)$. That is, we can only recover how much *extra* attention a person pays to each option as more of her friends select that option.

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 (4) 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 4 above.

Example 4 (continued). Let us keep all the structure of Example 4 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

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

The fact that $\mathrm{R}_1\left(1 \mid \{1\}\right)$ and $\mathrm{R}_1\left(1 \mid \{1, 2\}\right)$ 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:

$$\det \begin{pmatrix} 1-\mathrm{Q}_1(2 \mid y_1, 0) & \mathrm{Q}_1(2 \mid y_1, 0) \\ 1-\mathrm{Q}_1(2 \mid y_1, 1) & \mathrm{Q}_1(2 \mid y_1, 1) \end{pmatrix} = \mathrm{Q}_1(2 \mid y_1, 1) - \mathrm{Q}_1(2 \mid y_1, 0) > 0.$$

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.[7] This section offers a variant of the model that accommodates to this possibility.[8]

Let us assume there is no default option o, so that $\overline{\mathcal{Y}} = \mathcal{Y}$. 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.[9] Formally, the probability that Person a selects (at the moment of choosing) alternative $v \in \mathcal{Y}$ is given by

$$\mathrm{P}_a(v \mid \mathbf{y}) = \mathrm{Q}_a(v \mid \mathbf{y}) \prod\nolimits_{v' \in \mathcal{Y}, v' \succ_a v} (1 - \mathrm{Q}_a(v' \mid \mathbf{y})) + 1(v = y_a) \prod\nolimits_{v' \in \mathcal{Y}} (1 - \mathrm{Q}_a(v' \mid \mathbf{y}))\,. \tag{5}$$

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

The condition that $Y \geq 3$ is equivalent to requiring at least 2 non-default alternatives in the initial model. Thus, it is not restrictive. The identification proof for the network and the consideration probabilities is very similar to the proof of the initial model. However, since there is

[7] Also, see Horan (2019).

[8] For alternative ways to close the model see, for example, Barseghyan et al. (2021b).

[9] In our empirical application we use this version of our model.

no default option, identification of preferences is more involved.

5.3. Non-Dominated Default Option

In Section 2.3, we interpreted the default option as the possibility of not-to-buy-any-alternative. With this interpretation in mind, it is reasonable to think that people might differ regarding the ranking of the default option. We show in this section that this variant of the model is also identified with an additional condition.

Let us first note that, from a technical perspective, we can model the possibility that people differ regarding their ranking of the default option by letting the set of available alternatives for each Person a be a subset of $\mathcal{Y}$ that is not dominated by the default option. We can do so because any option outside this set will never be chosen by the person. Let us call this set $\mathcal{Y}_a$, where we are making it clear that it varies across people. The identification result is as follows.

Proposition 5.3. *Suppose that Assumptions A1-A3 are satisfied and for each $a \in \mathcal{A}$ and each $y \in \mathcal{Y}_a$ there exists $a' \in \mathcal{N}_a$ such that $y \in \mathcal{Y}_{a'}$. Then, the set of connections Γ, the profile of strict preferences $\succ$, and the attention mechanism Q are point identified from P.*

The extra condition simply requires each person to have, for each alternative that is not dominated by the default option, at least one friend for whom that alternative is also not dominated by the default option. This condition ensures enough variability in the consideration probabilities to recover the ranking of preferences of each person. The identification strategy is otherwise similar to the initial one. (Note that for each Person a we can only recover the set of friends —or members of her reference group— that share with the person at least one element of the set of alternatives that are not dominated by the default.)

5.4. 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

$$\prod_{v\in\mathcal{C}} \mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})\right) \prod_{v\notin\mathcal{C}} \left(1 - \mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})\right)\right).$$

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 $\eta_a(\cdot \mid \mathbf{y})$ be an attention index function from $2^{\mathcal{Y}}$ to the positive reals. The value $\eta_a(\mathcal{C} \mid \mathbf{y})$ 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. (2022) for further details). We define the probability of facing consideration set $\mathcal{C}$ as

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

In this model, the consideration set probabilities are as in Brady and Rehbeck (2016). Since η_a can be identified only up to scale, we normalize the attention-index for the empty set to be 1 (i.e., $\eta(\emptyset \mid \mathbf{y}) = 1$ for all $\mathbf{y} \in \overline{\mathcal{Y}}^A$).[10] 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

$$\mathrm{P}_a\left(v \mid \mathbf{y}\right) = \sum_{\mathcal{C}\in 2^{\mathcal{Y}}: v\in\mathcal{C}} \mathrm{R}_a\left(v \mid \mathcal{C}\right) \frac{\eta_a\left(\mathcal{C} \mid \mathbf{y}\right)}{\sum_{\mathcal{D}\subseteq\mathcal{Y}} \eta_a(\mathcal{D} \mid \mathbf{y})}.$$

The probability of selecting the default option o is just $\eta_a\left(\emptyset \mid \mathbf{y}\right) / \sum_{\mathcal{D}\subseteq\mathcal{Y}} \eta_a(\mathcal{D} \mid \mathbf{y})$. To better

[10]If one instead normalizes $\sum_{\mathcal{D}\subseteq\mathcal{Y}} \eta_a(\mathcal{D} \mid \mathbf{y}) = 1$, then we get the consideration rule of Aguiar (2017).

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

$$\frac{\eta_a(\mathcal{C} \cup \{v\} \mid \mathbf{y})}{\eta_a(\mathcal{C} \mid \mathbf{y})} = \frac{\mathrm{Q}_a(v \mid \mathbf{y})}{1 - \mathrm{Q}_a(v \mid \mathbf{y})}.$$

Thus,

$$\mathrm{Q}_a(v \mid \mathbf{y}) = \frac{\eta_a(\mathcal{C} \cup \{v\} \mid \mathbf{y})}{\eta_a(\mathcal{C} \cup \{v\} \mid \mathbf{y}) + \eta_a(\mathcal{C} \mid \mathbf{y})}.$$

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 $\mathbf{y} \in \overline{\mathcal{Y}}^A$, there exists $\mathcal{C} \in 2^{\mathcal{Y}}$ such that $v \succ_a v'$ for all $v' \in \mathcal{C}$ and $\eta_a\left(\mathcal{C} \cup \{v\} \mid \mathbf{y}\right) > 0$.

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

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

(ii) $a^* \in \mathcal{N}_a,\ y_{a^*} \in \mathcal{C}$ and $y^*_{a^*} \notin \mathcal{C} \implies \eta_a\left(\mathcal{C} \mid \mathbf{y}\right) \geq \eta_a\left(\mathcal{C} \mid \mathbf{y}^*\right)$ with strict inequality for some $\mathcal{C}$.

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

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

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

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

Condition (i) in Proposition 5.4 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. Empirical Application

The application we study is based on an experimental dataset that has been used to compare the effectiveness of various models that aim to predict the visual focus of attention of individuals within a group.[11] In the 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. The game was implemented in several rounds and lasted about 30 minutes. Five participants were seated around a circle. A tablet was placed in front of each player and it recorded the direction of

[11] See Bai et al. (2019).

the sight of the player every 1/3 of a second. Each player can focus on 5 points –4 other players and the tablet. The dataset contains only information about the exact point of focus of each player (i.e. we do not know whether the person of focus was speaking or not). To simplify the exposition, we let the choice set of each person to be looking to the left, to the tablet, or to the right (i.e., left, tablet, or right). Although the choice set is rather small, players on average make at least 1 choice every 1.76 seconds in our data. We believe that such high frequency of choice forces players to not consider all available options, thus justifying the use of our model. The raw frequency of choices in the data is displayed in Table 9.

Table 9 – Marginal Shares (%)

	Player 1	Player 2	Player 3	Player 4	Player 5
left	40	35	30	29	50
tablet	20	38	5	15	16
right	40	27	65	56	34

Notes: The sample size is 7875. Numbers were rounded.

Our structural approach allows us to separate two key forces in explaining the data: directional sight preferences of players and peer effects in gazing. The two estimated effects 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, everything else equal, people spend more time looking to the left side of the screen as compared to the right side.[12] This so-called left-to-right bias has been also documented by experimental studies.[13] Second, in social environments, it has been observed that people automatically redirect their visual attention by following the gaze orientation of other people, a phenomenon called gaze following.[14]

Before offering more details on the dataset and our estimates, let us briefly describe the main findings. Interestingly, despite considerable differences in the raw marginal shares of the three choices across players, as reflected in Table 9, the estimated preferences for directional sight of all players coincide. (Let us highlight here that in our estimation we allowed for full heterogeneity in

[12]See, for example, `https://www.nngroup.com/articles/horizontal-attention-original-research/`.

[13]See Spalek and Hammad (2004, 2005) and Reutskaja, Nagel, Camerer and Rangel (2011). See also Maass and Russo (2003) and reference therein.

[14]See, for example, Gallup, Hale, Sumpter, Garnier, Kacelnik, Krebs and Couzin (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`.

preferences across people.) Consistent with the left-to-right bias, all players preferences agree in that

$$\text{left} \succ \text{tablet} \succ \text{right}.$$

The estimates for peer effects in consideration sets (in the data) are positive and relevant. These findings emphasize the importance of peer effects on attention and provides indirect evidence of the left-to-right bias, which is important for advertisement and marketing.

We next present more details on the dataset we use and some assumptions we make to simplify the exposition. (See Appendix E for extra information.)

Data The dataset contains 7,875 observations of the direction of each of 5 player's sight. Although the frequency of sight measure was rather high (3 measurements per second), about 26 percent of the observations have more than one player changing the direction of sight between consecutive measurements. Thus, we treat the dataset 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, tablet, and right.[15] (The original dataset has five options as each person can look at one of her four opponents or her tablet.)

Consideration probabilities and choices of friends We assume that consideration probabilities on the direction of sight (i.e., left, tablet, or right) vary with the number of peers that look in the same direction. That is, each player is more likely to consider looking at a particular area if other players are doing so.[16]

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. This assumption is not necessary for our analysis and can be dropped. However, it decreases the computational burden

[15] 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, i.e., $3^5 = 243$ vs. $5^5 = 3125$.

[16] Allowing for dependence of consideration probabilities on the current selection does not change the estimated preferences. See Appendix E for further details.

and simplifies the exposition. Moreover, our estimation results suggest that this assumption holds. In particular, the estimated consideration probabilities increase from 3 to 4 peers, which is the maximal number of friends, for all players (see Figure 2).

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.

Estimation Since we work with discrete-time data (Dataset 2), we follow the following general estimation procedure. Let $\theta = (\Gamma, \succ, \mathrm{Q})$ be an element of the space of possible parameters we want to estimate. (We normalize the intensity parameter λ_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 Γ and $\succ$ is a finite set. For each θ 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 $\mathrm{L}_T(\theta) = \Sigma_{t=0}^{T-1} \ln \mathcal{P}_{\iota(\mathbf{y}_t), \iota(\mathbf{y}_{t+1})}(\theta, \Delta)$, where $\iota(\mathbf{y}) \in \left\{1, 2, \ldots, \overline{\mathcal{Y}}^A\right\}$ is the position of $\mathbf{y}$ according the lexicographic order, and $\mathcal{P}_{k,m}(\theta, \Delta)$ is the (k, m)-th element of the matrix $\mathcal{P}(\theta, \Delta)$. Finally, the maximum likelihood estimator of the true parameter value can be defined as[17]

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

Given that we assume that the network structure is known (the network is complete), we only need to estimate $\succ$ and Q.

[17]We assesses the finite-sample performance of the proposed estimator in Appendix B.

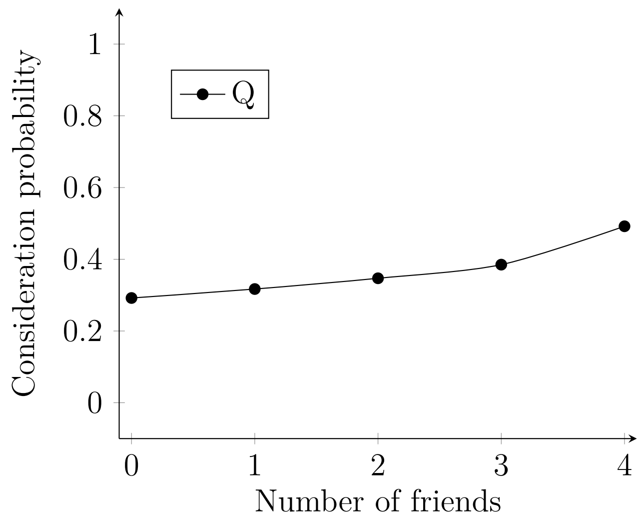


Figure 1 – Consideration probability as a function of the number of friends looking in the same direction. Model I(a).

We estimate a few specifications of the model. We present two of these specifications next and display the other ones in Appendix E.3. In Model I(a), we assume there is no heterogeneity in consideration probabilities across options and players, i.e., $\mathrm{Q}(\cdot) = \mathrm{Q}_a(v \mid \cdot)$ 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). Note that although monotonicity of Q in the number of friends was not imposed in the estimation, the estimated probabilities are indeed monotone increasing. The estimated preferences for directional sight coincide for all players. Specifically, all players prefer looking to the left, then to tablet, and then to the right. That is, we obtain that, for each $a \in \mathcal{A}$,

$$\text{left} \succ_a \text{table} \succ_a \text{right}.$$

Thus, in the first specification of our model, all the players show the left-to-right bias.

Figure 2 shows the estimates for the consideration probabilities for Model I(b). Similar to the case with homogeneity, many of the consideration probabilities for each direction of sight are indeed increasing in the number of players looking at that direction. This monotonicity also suggests that indeed every player is connected to all other players (e.g., if Player a had only 2 friends, then the

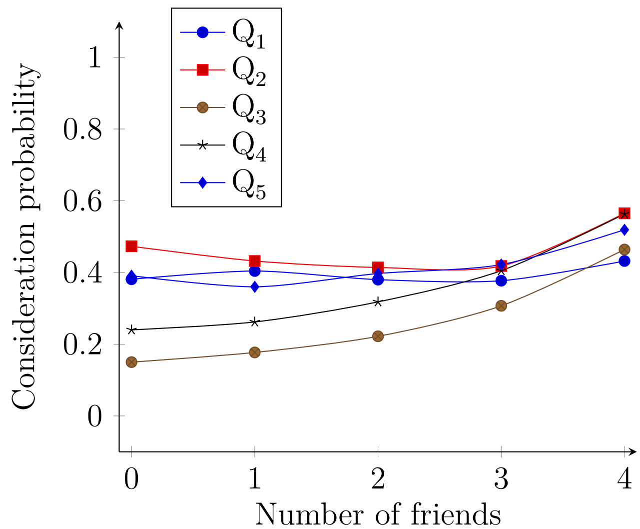


Figure 2 – Consideration probabilities for different players as functions of the number of friends looking in the same direction. Model I(b).

consideration probability would be the same for 2, 3, and 4 friends). The estimated preference orders coincide in Models I(a) and (b). 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 Table 9), the preferences for directional sight of all players coincide and are consistent with the left-to-right bias. As a robustness check, to address possible concerns with the dynamic of players' interactions in different stages of the game (e.g. learning), 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). The consideration probabilities are similar to the ones estimated using the whole sample, and the estimated preferences of all players are again left $\succ_a$ table $\succ_a$ right. (See Appendix E.2 for further details.)

7. 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. (Appendix C shows that the empirical predictions of our model differ from the ones of the model with peer effects in preferences.)

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 documented left-to-right bias in directional sighting and positive peers effects in gazing.

We believe that our approach to peer effects in consideration sets could be incorporated in various empirical studies of technology adoption and diffusion. For instance, Aral, Muchnik and Sundararajan (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 (2022) "Random utility and limited consideration," *Quantitative Economics*, Accepted.

Aguiar, Victor H ⓡ 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 (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.

Kashaev, Nail ⓡ Victor H. Aguiar (2022) "A random attention and utility model," *Journal of Economic Theory*, 204, 105487, https://doi.org/10.1016/j.jet.2022.105487.

Lazzati, Natalia (2020) "Co-Diffusion of Technologies in Social Networks," *American Economic Journal: Microeconomics*, 12, 193–228.

Lee, Sang Yup (2015) "Interpersonal influence on online game choices," *Computers in Human Behavior*, 45, 129–136.

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.

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 equilibrium μ exists and it is unique. Thus, we only need to prove that A1 implies that the Markov chain induced by our model is irreducible. First note that, under A1, for each $a \in \mathcal{A}$, $v \in \mathcal{Y}$, and $\mathbf{y} \in \overline{\mathcal{Y}}^A$, we have that

$$1 > \mathrm{P}_a\left(v \mid \mathbf{y}\right) = \mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v\left(\mathbf{y}\right)\right) \prod_{v' \in \mathcal{Y}, v' \succ_a v} \left(1 - \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}\left(\mathbf{y}\right)\right)\right) > 0.$$

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

Proof of Proposition 3.2: Let v_a^* be the most preferred alternative of Person a according to $\succ_a$. Given that $\mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v\left(\mathbf{y}\right)\right) = \mathrm{Q}_a\left(v \mid \mathrm{N}_a^v\left(\mathbf{y}\right)\right)$ for all $v \in \mathcal{Y}$, and $a \notin \mathcal{N}_{a'}$ for all $a' \in \mathcal{A}$, we have that

$$\mu_a\left(v_a^*\right) = \sum_{\mathbf{y}_{-a} \in \overline{\mathcal{Y}}^{A-1}} \mathrm{Q}_a\left(v_a^* \mid \mathrm{N}_a^v\left(\mathbf{y}\right)\right) \mu_{-a}\left(\mathbf{y}_{-a}\right)$$

.

If $\mathcal{N}'_a \subseteq \mathcal{N}_a$, then $\mathrm{Q}_a\left(v_a^* \mid \mathrm{N}_a^v\left(\mathbf{y}\right)\right) \geq \mathrm{Q'}_a\left(v_a^* \mid \mathrm{N}_a^v\left(\mathbf{y}\right)\right)$. The result follows because, by A3, $\mathrm{Q}_a\left(v \mid \mathrm{N}_a^v\left(\mathbf{y}\right)\right)$ is weakly increasing in $\mathrm{N}_a^v\left(\mathbf{y}\right)$. ■

Proof of Proposition 4.1: By A1, $\mathrm{P}_a\left(\cdot \mid \mathbf{y}\right)$ has full support for each $\mathbf{y} \in \overline{\mathcal{Y}}^A$. By A2 and A3, $\mathrm{P}_a\left(v \mid \mathbf{y}\right)$ is decreasing in $\mathrm{N}_a^{v'}\left(\mathbf{y}\right)$ for each $v' \succ_a v$ (with strict inequality at zero). 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 $\mathrm{N}_a^{v'}\left(\mathbf{y}\right)$ for each $v' \neq v$, we can recover Person a's upper level set that corresponds to option v. That is,

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

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

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

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

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

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

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

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

Then we can also recover λ_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 $\left(\mathrm{P}_a\right)_{a\in\mathcal{A}}$ from $\mathcal{M}$ as in the proof of Proposition 4.2 ■

Proof of Proposition 5.1: Note that expression (4) can be rewritten as follows

$$\mathrm{P}_a\left(v \mid \mathbf{y}\right) = \sum_{\mathcal{C}\subseteq\mathcal{Y}} \mathrm{R}_a\left(v \mid \mathcal{C}\right) \prod_{v'\in\mathcal{C}} \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right) \prod_{v''\notin\mathcal{C}} \left(1 - \mathrm{Q}_a\left(v'' \mid y_a, \mathrm{N}_a^{v''}(\mathbf{y})\right)\right) =$$

$$\mathrm{Q}_a\left(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})\right) \sum_{\mathcal{C} \subseteq \mathcal{Y}, v \in \mathcal{C}} \mathrm{R}_a(v \mid \mathcal{C}) \prod_{v' \in \mathcal{C} \setminus \{v\}} \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}_a^{v'}(\mathbf{y})\right) \prod_{v'' \notin \mathcal{C}}\left(1-\mathrm{Q}_a\left(v'' \mid y_a, \mathrm{N}_a^{v''}(\mathbf{y})\right)\right).$$

Thus, by A2 and A3, we can state whether $a' \in \mathcal{N}_a$ by checking whether $\mathrm{P}_a(v \mid y_1 = o, \ldots, y_A = o)$ 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 $\mathrm{N}_a^v(\mathbf{y}) = 0$ 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

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

Note that

$$\mathrm{P}_a(v \mid \mathbf{y}') / \mathrm{P}_a(v \mid \mathbf{y}) = \mathrm{Q}_a(v \mid y_a, 1) / \mathrm{Q}_a(v \mid y_a, 0).$$

Also

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

Thus, by A3, $\mathrm{Q}_a(v \mid y_a, 0)$ and $\mathrm{Q}_a(v \mid y_a, 1)$ can be recovered from the data. By implementing a similar procedure for different values of $\mathrm{N}_a^v(\mathbf{y})$ 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 $(\mathrm{Q}_a)_{a \in \mathcal{A}}$.

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 $\mathrm{R}_a(1 \mid \mathcal{C})$ for all $\mathcal{C}$. To simplify the exposition we will write

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

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

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

To present the ideas more clear let $A(N, Y)$ be the matrix of coefficients in front of the $\mathrm{R}_a^{'}$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

$$\prod_{k\in\mathcal{Y}} \mathrm{Q}^{\mathbb{1}(k\in\mathcal{C})}\left(k \mid \mathrm{N}^k\right)$$

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

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

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

Proof.

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

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

and

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

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

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

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

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

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

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

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

has full column rank since $\det(A(1,2)) = \mathrm{Q}^1(2 \mid 1) - \mathrm{Q}^1(2 \mid 0) > 0$. 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. ■

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 $\mathrm{P}_{a_1}(v \mid \mathbf{y}) = \mathrm{P}_{a_1}(v \mid \mathbf{y}')$ 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 $y_{a_1} \neq v^*_{a_1}$

$$\mathrm{P}_{a_1}\left(v^*_{a_1} \mid \mathbf{y}\right) = \mathrm{Q}_{a_1}\left(v^*_{a_1} \mid \mathbf{y}\right)$$

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 $\mathrm{P}_{a_1}(v \mid \mathbf{y})$ 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 Γ. 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.) Recall that for the probability of picking the best alternative of Person a_1, $v^*_{a_1}$, is

$$\mathrm{P}_{a_1}(v^*_{a_1} \mid \mathbf{y}) = \mathrm{Q}_{a_1}\left(v^*_{a_1} \mid y_{a_1}, \mathrm{N}^{v^*_{a_1}}_{a_1}(\mathbf{y})\right)$$

for any $y_{a_1} \neq v^*_{a_1}$. Hence, if any of friends of a_1 switches from the second best ($v^{**}_{a_1}$) to the third best ($v^{***}_{a_1}$) option or from $v^{***}_{a_1}$ to $v^{**}_{a_1}$, then $\mathrm{P}_{a_1}(v^*_{a_1} \mid \mathbf{y})$ will not change.

Similarly, the probability of picking the second best is

$$\mathrm{P}_{a_1}(v^{**}_{a_1} \mid \mathbf{y}) = \mathrm{Q}_{a_1}\left(v^{**}_{a_1} \mid y_{a_1}, \mathrm{N}^{v^{**}_{a_1}}_{a_1}(\mathbf{y})\right)\left(1 - \mathrm{Q}_{a_1}\left(v^{*}_{a_1} \mid y_{a_1}, \mathrm{N}^{v^{*}_{a_1}}_{a_1}(\mathbf{y})\right)\right)$$

for any $\mathbf{y}$ such that $y_{a_1} \neq v^{**}_{a_1}$ and $\mathrm{N}^{v^{*}_{a_1}}_{a_1}(\mathbf{y}) \in \{0,1\}$. Thus, the friend changing from $v^{*}_{a_1}$ ($v^{***}_{a_1}$) to $v^{***}_{a_1}$ ($v^{*}_{a_1}$) will increase (decrease) $\mathrm{P}_{a_1}(v^{**}_{a_1} \mid \mathbf{y})$.

Finally, the probability of picking the last option is

$$\mathrm{P}_{a_1}(v^{***}_{a_1} \mid \mathbf{y}) = \mathrm{Q}_{a_1}\left(v^{***}_{a_1} \mid y_{a_1}, \mathrm{N}^{v^{***}_{a_1}}_{a_1}(\mathbf{y})\right)\left(1 - \mathrm{Q}_{a_1}\left(v^{*}_{a_1} \mid y_{a_1}, \mathrm{N}^{v^{*}_{a_1}}_{a_1}(\mathbf{y})\right)\right)\left(1 - \mathrm{Q}_{a_1}\left(v^{**}_{a_1} \mid y_{a_1}, \mathrm{N}^{v^{**}_{a_1}}_{a_1}(\mathbf{y})\right)\right).$$

The changes in this probability because of switches of a friend from $v^{*}_{a_1}$ to $v^{**}_{a_1}$ or from $v^{**}_{a_1}$ to $v^{*}_{a_1}$ are ambiguous: $\mathrm{P}_{a_1}(v^{***}_{a_1} \mid \mathbf{y})$ may increase, decrease, or stay constant.

Next take option 1 and consider changes in the probability of picking it by Person 1 when her friend of switches from option 2 to option 3 and from option 3 to option 2. Then repeat the same exercise for option 2 (switches from 1 to 3 and from 3 to 1) and option 3 (switches from 1 to 3 and from 3 to 1). Note that one of these probabilities, say for option 1, will be invariant to switches, another one will increase and decrease (say option 2). The third probability (for option 3) may behave differently.

Case 1. The third probability is increasing with one switch and decreasing with the opposite one. Then we can conclude with certainty that option 1 is the first best option. Hence we can identify

$$\mathrm{Q}_{a_1}\left(v^{*}_{a_1} \mid y_{a_1}, \mathrm{N}^{v^{*}_{a_1}}_{a_1}(\mathbf{y})\right) = \mathrm{P}_{a_1}(v^{*}_{a_1} \mid \mathbf{y})$$

for all $y_{a_1} \neq v^{*}_{a_1} = 1$. Next note that

$$\frac{\mathrm{P}_{a_1}(2 \mid \mathbf{y})}{1 - \mathrm{P}_{a_1}(1 \mid \mathbf{y})} = \begin{cases} \mathrm{Q}_a(2 \mid y_{a_1}, \mathrm{N}^2_{a_1}(\mathbf{y})), & \text{if } 2 \succ_{a_1} 3 \\ \mathrm{Q}_a(2 \mid y_{a_1}, \mathrm{N}^2_{a_1}(\mathbf{y}))(1 - \mathrm{Q}_a(3 \mid y_{a_1}, \mathrm{N}^3_{a_1}(\mathbf{y}))), & \text{if } 3 \succ_{a_1} 2 \end{cases}$$

for any $y_{a_1} \notin \{1,2\}$. Thus a switch from option 1 to option 3 (or from option 3 to option 1) should identify the second best and the third best options.

Case 2. The third probability is invariant to changes. Then we can conclude with certainty that option 2 is the second best option. We also identify the fist best the the third best options since $\mathrm{P}_{a_1}(v^{**}_{a_1} \mid \mathbf{y})$ increases when a friend switches from the first best to the third best option and increases when the switch is opposite.

Case 3. Non of the above two cases. Then we can conclude with certainty that option 1 is the first best, option 2 is the second best, and option 3 is the third best option.

Thus we can identify preferences of a_1. Since the choice of a_1 was arbitrary we can identify $\succ = (\succ_a)_{a\in\mathcal{A}}$.

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 $\mathrm{Q}_{a_1}(y^*_{a_1} \mid y_{a_1}, \mathrm{N}^{y^*_{a_1}}_{a_1}(\mathbf{y}))$ for any $\mathbf{y}$ such that $y^*_{a_1} \neq y_{a_1}$ since

$$\mathrm{Q}_{a_1}\left(y^*_{a_1} \mid y_{a_1}, \mathrm{N}^{y^*_{a_1}}_{a_1}(\mathbf{y})\right) = \mathrm{P}_{a_1}(y^*_{a_1} \mid \mathbf{y}).$$

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

$$\prod\nolimits_{v'\neq y_{a_1}} \left(1 - \mathrm{Q}_a\left(v' \mid y_{a_1}, \mathrm{N}^{v'}_{a_1}(\mathbf{y})\right)\right)$$

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

$$\mathrm{Q}_{a_1}(y^*_{a_1} \mid y_{a_1}, \mathrm{N}^{y^*_{a_1}}_{a_1}(\mathbf{y})) = \frac{\mathrm{P}_{a_1}(y^*_{a_1} \mid \mathbf{y}) - \prod_{v'\neq y^*_{a_1}} \left(1 - \mathrm{Q}_a\left(v' \mid y^*_{a_1}, \mathrm{N}^{v'}_{a_1}(\mathbf{y})\right)\right)}{1 - \prod_{v'\neq y^*_{a_1}} \left(1 - \mathrm{Q}_a\left(v' \mid y^*_{a_1}, \mathrm{N}^{v'}_{a_1}(\mathbf{y})\right)\right)}.$$

Hence, we can identify $\mathrm{Q}_{a_1}(y^*_{a_1} \mid y_{a_1}, \mathrm{N}^{y^*_{a_1}}_{a_1}(\mathbf{y}))$ 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 $\mathrm{Q}_{a_1}(y^{**}_{a_1} \mid y_{a_1}, \mathrm{N}^{y^{**}_{a_1}}_{a_1}(\mathbf{y}))$ since

$$\mathrm{Q}_{a_1}(y^{**}_{a_1} \mid y_{a_1}, N^{y^{**}_{a_1}}_{a_1}(\mathbf{y})) = \frac{\mathrm{P}_{a_1}(y^{**}_{a_1} \mid \mathbf{y}) - \prod_{v'\neq y^{**}_{a_1}} \left(1 - \mathrm{Q}_{a_1}\left(v' \mid y^{**}_{a_1}, \mathrm{N}^{v'}_{a_1}(\mathbf{y})\right)\right)}{1 - \mathrm{Q}_{a_1}(y^*_{a_1} \mid y_{a_1}, \mathrm{N}^{v'}_{a_1}(\mathbf{y})) - \prod_{v'\neq y^{**}_{a_1}} \left(1 - \mathrm{Q}_{a_1}\left(v' \mid y^{**}_{a_1}, \mathrm{N}^{v'}_{a_1}(\mathbf{y})\right)\right)},$$

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

Proof of Proposition 5.3: By A1, $\mathrm{P}_a(\cdot \mid \mathbf{y})$ has full support on $\overline{\mathcal{Y}}_a = \mathcal{Y}_a \cup \{o\}$ for each $\mathbf{y} \in \times_{a\in\mathcal{A}}\overline{\mathcal{Y}}_a$ and for every a. By A2 and A3, $\mathrm{P}_a(v \mid \mathbf{y})$ is weakly decreasing in $\mathrm{N}_a^{v'}(\mathbf{y})$ for each $v' \succ_a v$. Since $\mathrm{N}_a^{v'}(\mathbf{y})$ can take at lest two values (every option is in the choices set of at least one friend), 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 $\mathrm{N}_a^{v'}(\mathbf{y})$ for each $v' \neq v$, we can recover

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

By repeating the exercise with each alternative, we can recover $\succ_a$ over $\mathcal{Y}_a$. Finally, suppose that y_a^* is the most preferred alternative for Person a. Then,

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

It follows that we can recover $\mathrm{Q}_a\left(y_a^* \mid y_a, \mathrm{N}_a^{y_a^*}(\mathbf{y})\right)$ for each $\mathbf{y} \in \times_{a\in\mathcal{A}}\overline{\mathcal{Y}}_a$. By proceeding in descending preference ordering we can then recover $\mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^{v}(\mathbf{y}))$ for all $v \in \mathcal{Y}_a$ (and each $\mathbf{y} \in \times_{a\in\mathcal{A}}\overline{\mathcal{Y}}_a$). ∎

Proof of Proposition 5.4: 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 $\mathrm{P}_{a_1}(v \mid \mathbf{y}) = \mathrm{P}_{a_1}(v \mid \mathbf{y}')$ 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 $\mathbf{y}$

$$\frac{\mathrm{P}_{a_1}\left(v_{a_1}^* \mid \mathbf{y}\right)}{\mathrm{P}_{a_1}(o \mid \mathbf{y})} = \sum\nolimits_{\mathcal{C}\in 2^{\mathcal{Y}}:\, v_{a_1}^*\in\mathcal{C}} \eta_{a_1}(\mathcal{C} \mid \mathbf{y})$$

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 $\mathrm{P}_{a_1}(v \mid \mathbf{y}) / \mathrm{P}_{a_1}(o \mid \mathbf{y})$ 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 Γ. 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 $\mathbf{y}^* = (o, o, \ldots, o)'$. Note that

$$\frac{\mathrm{P}_{a_1}\left(v^*_{a_1} \mid \mathbf{y}^*\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}^*\right)} = \sum\nolimits_{\mathcal{C}\in 2^{\mathcal{Y}}:\, v^*_{a_1}\in\mathcal{C}} \eta_{a_1}\left(\mathcal{C} \mid \mathbf{y}^*\right)$$

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

$$\frac{\mathrm{P}_{a_1}\left(v^{**}_{a_1} \mid \mathbf{y}^*\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}^*\right)} = \sum\nolimits_{\mathcal{C}\in 2^{\mathcal{Y}}:\, v^*_{a_1}\notin\mathcal{C},\, v^{**}_{a_1}\in\mathcal{C}} \eta_{a_1}\left(\mathcal{C} \mid \mathbf{y}^*\right)$$

will increase if any of friends of a_1 switches to an element of $\mathcal{C}$ that corresponds to a strict inequality that and is different from $v^*_{a_1}$. 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 $Y \succ_{a_1} Y-1 \succ_{a_1} Y-2 \succ_{a_1} \cdots \succ_{a_1} 2 \succ_{a_1} 1$. Note that we have the following system of Y equations

$$\frac{\mathrm{P}_{a_1}\left(k \mid \mathbf{y}\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}\right)} = \begin{cases} \eta_{a_1}(\{1\} \mid \mathbf{y}), & k = 1, \\ \sum_{\mathcal{C}\subseteq\{1,\ldots,k-1\}} \eta_{a_1}(\mathcal{C}\cup\{k\} \mid \mathbf{y}), & k = 2, \ldots, Y. \end{cases}$$

Note that η_{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, η_{a_1} can not be identified without more restrictions. Suppose η_{a_1} is multiplicative. So we only need to identify $\eta_{a_1}(\{k\} \mid \mathbf{y})$ for all $k \in \mathcal{Y}$. Then, first, we can identify $\eta_{a_1}(\{1\} \mid \mathbf{y})$ since

$$\frac{\mathrm{P}_{a_1}\left(1 \mid \mathbf{y}\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}\right)} = \eta_{a_1}(\{1\} \mid \mathbf{y}).$$

Next,

$$\frac{\mathrm{P}_{a_1}(2 \mid \mathbf{y})}{\mathrm{P}_{a_1}(o \mid \mathbf{y})} = \eta_{a_1}(\{2\} \mid \mathbf{y}) + \eta_{a_1}(\{1,2\} \mid \mathbf{y}) = \eta_{a_1}(\{2\} \mid \mathbf{y}) + \eta_{a_1}(\{1\} \mid \mathbf{y})\eta_{a_1}(\{2\} \mid \mathbf{y}).$$

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

$$\eta_{a_1}(\{k\} \mid \mathbf{y}) = \frac{\mathrm{P}_{a_1}(k \mid \mathbf{y})}{\mathrm{P}_{a_1}(o \mid \mathbf{y})} \cdot \frac{1}{\sum_{\mathcal{C} \subseteq \{1,\ldots,k-1\}} \eta_a(\mathcal{C} \mid \mathbf{y})}.$$

The same argument can be applied if η_{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., $\eta_{a_1}(\{1,2\} \mid \mathbf{y}) = \phi_{\{1,2\}}(\eta_{a_1}(\{1\} \mid \mathbf{y}), \eta_{a_1}(\{2\} \mid \mathbf{y}))$, where $\phi_{1,2}$ is strictly increasing in both arguments).

Suppose that η_{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 $\eta_{a_1}(\{2\} \mid \mathbf{y})$. Hence, using the second equation we identify $\eta_{a_1}(\{1,2\} \mid \mathbf{y})$ and, thus, attention indexes for all sets of cardinality 2. Repeating the above arguments recursively we can identify η_{a_1}.

It is left to show that if η_{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 $\eta_{a_1}(\{2\} \mid \mathbf{y}) = \eta_{a_1}(\{1,2\} \mid \mathbf{y})$, from the second equation we identify

$$\eta_{a_1}(\{2\} \mid \mathbf{y}) = \eta_{a_1}(\{1,2\} \mid \mathbf{y}) = \frac{\mathrm{P}_{a_1}(2 \mid \mathbf{y})}{2\,\mathrm{P}_{a_1}(o \mid \mathbf{y})}.$$

Repeating the above arguments for every equation we get that

$$\eta_{a_1}(\mathcal{C} \mid \mathbf{y}) = \frac{\mathrm{P}_{a_1}\left(v^*_{a_1,\mathcal{C}} \mid \mathbf{y}\right)}{v^*_{a_1,\mathcal{C}}\,\mathrm{P}_{a_1}(o \mid \mathbf{y})},$$

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 η_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

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

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

$$\mathrm{Q}_a(v \mid \mathbf{y}) = \frac{\eta_a(\{y\} \mid \mathbf{y})}{1 + \eta_a(\{v\} \mid \mathbf{y})}.$$

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

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

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

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

Hence, by induction

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

As a result,

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

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

$$\prod\nolimits_{y\in A} \mathrm{Q}_a(y \mid \mathbf{y}) \prod\nolimits_{y'\in\mathcal{Y}\setminus A}(1-\mathrm{Q}_a(y' \mid \mathbf{y}))$$

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

$$\prod\nolimits_{y\in A} \mathrm{Q}_a(y \mid \mathbf{y}) \prod\nolimits_{y'\in\mathcal{Y}\setminus A}(1-\mathrm{Q}_a(y' \mid \mathbf{y})) = \frac{\eta_a(A \mid \mathbf{y})}{\sum_{C\subseteq\mathcal{Y}} \eta_a(C \mid \mathbf{y})}.$$

■

B. Finite Sample Performance of the Estimator

In this appendix we evaluate the performance of our estimator by means of simulations.

Recall that $\theta = (\Gamma, \succ, \mathrm{Q})$ is an element of the space of possible parameters we want to estimate. Since the number of people in the network and the number of choices is finite, the parameter space for Γ and $\succ$ is a finite set. For each θ we construct the transition rate matrix $\mathcal{M}(\theta)$ using Equations (1) and (2). The transition matrix then is

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

The log-likelihood function is $\mathrm{L}_T(\theta) = \Sigma_{t=0}^{T-1} \ln \mathcal{P}_{\iota(\mathbf{y}_t),\iota(\mathbf{y}_{t+1})}(\theta, \Delta)$, where $\iota(\mathbf{y}) \in \left\{1, 2, \ldots, \overline{\mathcal{Y}}^A\right\}$ is the position of $\mathbf{y}$ according the lexicographic order, and $\mathcal{P}_{k,m}(\theta, \Delta)$ is the (k, m)-th element of the matrix $\mathcal{P}(\theta, \Delta)$. Finally, the maximum likelihood estimator of the true parameter value can be defined as

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

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

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

$$\mathrm{Q}\left(v \mid 0\right) = \frac{1}{4}, \quad \mathrm{Q}\left(v \mid 1\right) = \frac{3}{4}, \quad \mathrm{Q}\left(v \mid 2\right) = \frac{7}{8},$$

and

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

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

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

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

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

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

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

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

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

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

		Sample Size					
Attention Probabilities		10	50	100	500	1000	5000
$Q(v \mid 0)$	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
$Q(v \mid 1)$	Bias	66.4	53	46.9	27.2	18	2.7
	RMSE	103.9	66.9	55.6	31	20.9	5.6
$Q(v \mid 2)$	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 Δ equal to 2500, 500, 250, 50, 25, and 5. The number of replications is 1000.

simulations are presented in Table 10. 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.

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 Γ 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 still 32). The experiment was conducted 500 times for different sample sizes. Table 11 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 11 – 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 Δ equal to 2500, 500, 250, and 50. The number of replications is 500.

C. Relation with a Peer-effects-in-preferences Model

In this appendix, we consider a more classical model of social interactions where choices of peers affects ones preferences but not consideration and consideration is full (i.e., all alternatives are always considered) and compare it with the extension of our initial model presented in Section 5.4.

To build a model of peer effects in preferences, we follow Brock and Durlauf (2001) and Brock and Durlauf (2002) and assume that the utility that Person a gets from option v given configuration $\mathbf{y}$ is

$$u_a^v + S_a^v\left(y_a, \mathrm{N}_a^v\left(\mathbf{y}\right)\right) + \varepsilon_a^v,$$

where u_a^v is the private utility of Person a associated with option v; $S_a^v\left(y_a, \mathrm{N}_a^v\left(\mathbf{y}\right)\right)$ is strictly increasing in $\mathrm{N}_a^v\left(\mathbf{y}\right)$ the social utility associated with the choice; and ε_a^v is a random utility term, independently and identically distributed across persons with a Type I extreme-value distribution. Given the assumptions about the distribution of ε_a^v and the assumption that consideration is full, the distribution over options conditional on the network configuration can be represented by the standard logit formula:

$$\mathrm{P}_a\left(v \mid \mathbf{y}\right) = \frac{\exp\left(u_a^v + S_a^v\left(y_a, \mathrm{N}_a^v\left(\mathbf{y}\right)\right)\right)}{\sum_{v'} \exp\left(u_a^{v'} + S_a^{v'}\left(y_a, \mathrm{N}_a^{v'}\left(\mathbf{y}\right)\right)\right)}. \tag{6}$$

Recall that in our model

$$\mathrm{P}_a\left(v \mid \mathbf{y}\right) = \textstyle\sum_{\mathcal{C} \in 2^{\mathcal{Y}}: v \in \mathcal{C}} \mathbb{1}\left(\, v \succ_a v', v' \in \mathcal{C}\,\right) \frac{\eta_a\left(\mathcal{C} \mid \mathbf{y}\right)}{\sum_{\mathcal{D} \subseteq \mathcal{Y}} \eta_a(\mathcal{D} \mid \mathbf{y})}. \tag{7}$$

The following result shows that our model and the above model of peer effects in preference have different empirical implications.

Proposition C.1. *Any* P *can be consistent with the model represented by (6) or with the model represented by (7), but not both.*

Proof. By way of contradiction assume that some P is consistent with both (6) and (7). Since, it is consistent with (7), by Proposition 5.4 we can identify Γ and $\succ$. Fix any Person a_1 and

$\mathbf{y}^* = (o, o, \ldots, o)\prime$. Let $v^*_{a_1}$ and $v^{**}_{a_1}$ be the first and the second best preferred alternatives of a_1. Then

$$\frac{\mathrm{P}_{a_1}\left(v^{**}_{a_1} \mid \mathbf{y}^*\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}^*\right)} = \sum\nolimits_{\mathcal{C}\in 2^{\mathcal{Y}}:\, v^*_{a_1}\notin\mathcal{C},\, v^{**}_{a_1}\in\mathcal{C}} \eta_{a_1}\left(\mathcal{C} \mid \mathbf{y}^*\right)$$

will not change if any of friends of a_1 switches to $v^*_{a_1}$. However, since P is consistent with (6), then

$$\frac{\mathrm{P}_{a_1}\left(v^{**}_{a_1} \mid \mathbf{y}^*\right)}{\mathrm{P}_{a_1}\left(o \mid \mathbf{y}^*\right)} = \frac{\exp\left(u^{v^{**}_{a_1}}_{a_1} + S^{v^{**}_{a_1}}_{a_1}\left(y_{a_1}, 0\right)\right)}{\exp\left(u^{o}_{a_1} + S^{o}_{a_1}\left(y_{a_1}, \left|\mathcal{N}_{a_1}\right|\right)\right)}$$

will increase if any of friends of a_1 switches to $v^*_{a_1}$. The contradiction completes the proof. ■

D. Gibbs Random Field Model

In this appendix, 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 $\left(\mathrm{P}_a\right)_{a\in\mathcal{A}}$, with a generic element given by

$$\mathrm{P}_a\left(v \mid \mathbf{y}\right) = \mathrm{Q}_a\left(v \mid y_a, \mathrm{N}^v_a\left(\mathbf{y}\right)\right) \prod\nolimits_{v'\in\mathcal{Y}, v'\succ_a v} \left(1 - \mathrm{Q}_a\left(v' \mid y_a, \mathrm{N}^{v'}_a\left(\mathbf{y}\right)\right)\right) \text{ for } v \in \mathcal{Y} \text{ and } \mathbf{y} \in \overline{\mathcal{Y}}^A.$$

A Gibbs equilibrium is defined as a joint distribution over the vector of choices $\mathbf{y}$, $\mathrm{P}\left(\mathbf{y}\right)$, that is able to generate $\left(\mathrm{P}_a\right)_{a\in\mathcal{A}}$ 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 $\mathrm{Q}_a\left(v \mid y_a, \mathrm{N}^v_a\left(\mathbf{y}\right)\right)$ 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 $\mathbf{y} \in \overline{\mathcal{Y}}^A$, $\mathrm{Q}_a(v \mid \mathbf{y}) \equiv \mathrm{Q}_a(v \mid \mathrm{N}_a^v(\mathbf{y}))$.

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

Proposition D.1. *If A1 and G1 are satisfied, then there exists a unique μ. Also, μ satisfies*

$$\mu(\mathbf{y}) = \frac{1}{\sum_{a\in\mathcal{A}} \lambda_a} \sum_{a\in\mathcal{A}} \lambda_a \mathrm{P}_a(y_a \mid \mathbf{y})\, \mu_{-a}(\mathbf{y}_{-a}) \ \textit{for each } \mathbf{y} \in \overline{\mathcal{Y}}^A.$$

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

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

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

In what follows, we will normalize the lambdas to the value of 1. From Proposition D.1, μ satisfies

$$\mu(\mathbf{y}) = \frac{1}{A}\sum_{a\in\mathcal{A}} \mathrm{P}_a(y_a \mid \mathbf{y})\,\mu_{-a}(\mathbf{y}_{-a}) \ \text{ for each } \mathbf{y} \in\overline{\mathcal{Y}}^A. \tag{8}$$

We next show that if $(\mathrm{P}_a)_{a\in A}$ is a set of compatible conditional distributions, then $\mu = \mathrm{P}$ solves (8). If we let $\mu = \mathrm{P}$, then right hand side of (8) is

$$\frac{1}{A}\sum_{a\in\mathcal{A}} \mathrm{P}_a(y_a \mid \mathbf{y})\sum_{v\in\overline{\mathcal{Y}}} \mathrm{P}(v,\mathbf{y}_{-a}) = \frac{1}{A}\sum_{a\in\mathcal{A}} \mathrm{P}(\mathbf{y}) = \frac{A}{A}\mathrm{P}(\mathbf{y}) = \mathrm{P}(\mathbf{y}).$$

Also, the left hand side of (8) is

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

Thus $\mu(\mathbf{y}) = \mathrm{P}(\mathbf{y})$ solves (8) for each $\mathbf{y} \in \overline{\mathcal{Y}}^A$.

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

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

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

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

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

The last result states that, under specific conditions, the Gibbs equilibrium coincides with μ. A similar connection is discussed in Blume and Durlauf (2003). The next proposition puts together all our previous ideas.

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

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 μ of the dynamic revision process coincides with the Gibbs equilibrium of the model. To see this notice that from our previous results we get

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

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

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

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

E. Additional Estimation Results

E.1. Additional Experiment

The dataset in Bai et al. (2019) contains two experiments with five players. We presented the results for one of them. We next state that the estimates for the other one are quire similar.[18] This experiment contains 4248 observations. Similar to the experiment in the main text (the Main Experiment), we present marginal shares of different alternatives in the Additional Experiment in Table 12.

Looking at Table 12 we see that, in the Additional Experiment, 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

[18] The full 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.

Table 12 – Marginal Shares in the Additional Experiment (%)

	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.

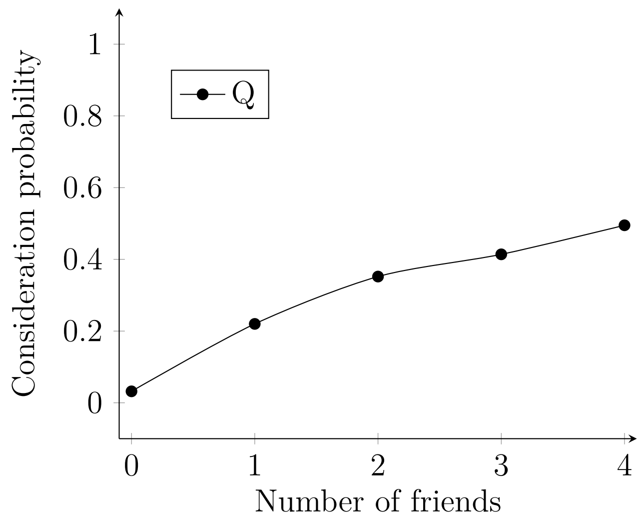


Figure 3 – Consideration probability as a function of the number of friends looking in the same direction. Model I(a). Additional Experiment.

to the right. Recall, that the Main Experiment has more balanced shares. In other words, 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.

We estimated the same specification. Recall that in Model I(a), we assume there is no heterogeneity in consideration probabilities across options and players, i.e., $Q(\cdot) = Q_a(v \mid \cdot)$ 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 3 shows the estimates for the consideration probabilities of Model I(a) for the Additional Experiment. Identically to the Main Experiment, the estimated preferences for directional sight coincide for all players. Specifically, all players prefer looking to the left, then to tablet, and then to the right. Thus, in the first specification of our model, all the players, as in the Main Experiment, show the left-to-right bias.

Figure 4 shows the estimates for the consideration probabilities for Model I(b) in Experiments

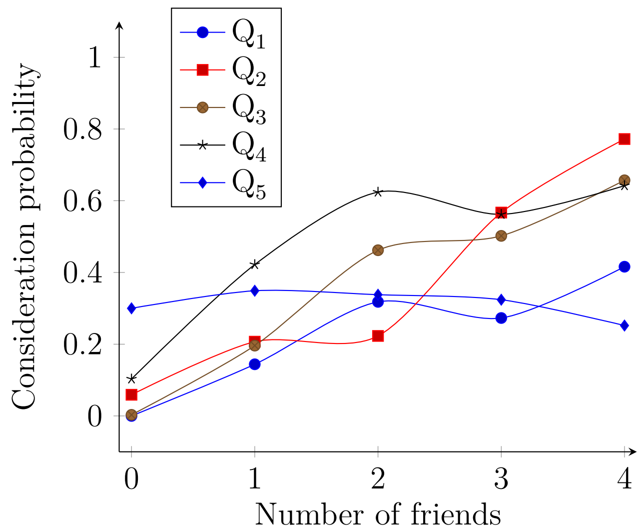


Figure 4 – Consideration probabilities for different players as functions of the number of friends looking in the same direction. Model I(b). Additional Experiment.

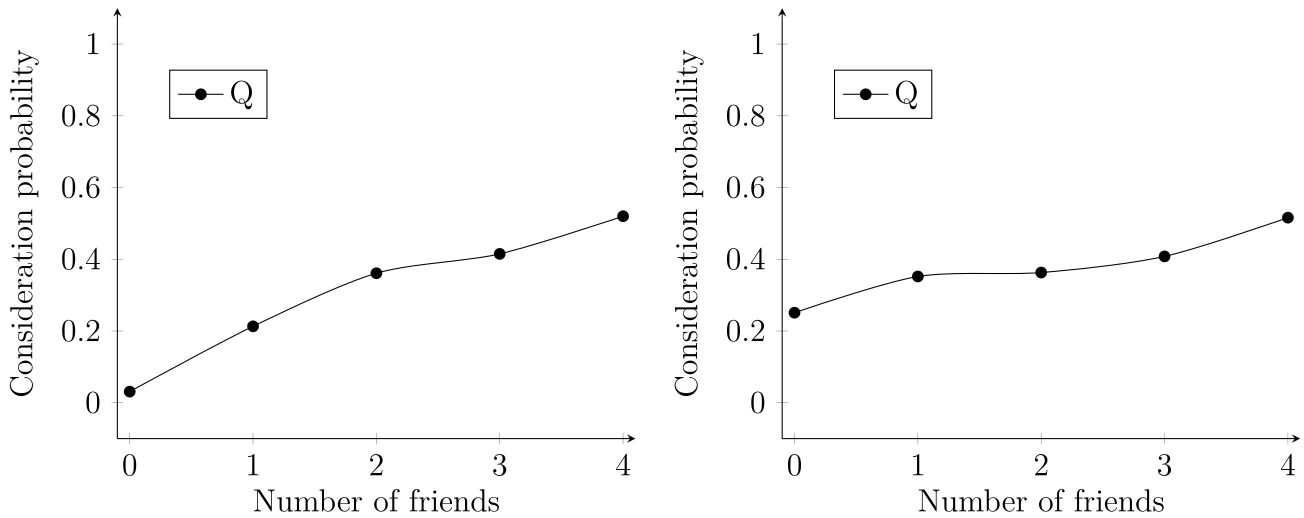
1 and 2. Similar to the case with homogeneity, many of the consideration probabilities for each direction of sight are indeed increasing in the number of players looking at that direction. The estimated preference orders coincide in Models I(a) and (b), for the two experiments. That is, here again, all 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 across players and across experiments (see Tables 12 and 9), the preferences for directional sight of all players in both experiments coincide.

E.2. 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 6 using two different subsamples. The first subsample consists of the observations coming from the middle half (i.e., we trowed 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 5-8. The estimated preference orders are presented in Tables 13-14.



(a) Additional Experiment **(b)** Main Experiment

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

Table 13 – Preferences. Model I(a).

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

E.3. Heterogeneity in Consideration Probabilities Across Choices

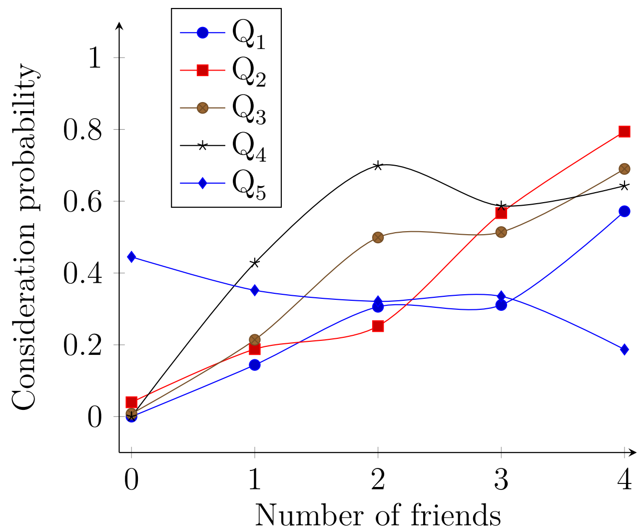
In this section we present results of estimation of the following specifications of the model:

(i) Model Ic: $\mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})) = \mathrm{Q}(v \mid \mathrm{N}_a^v(\mathbf{y}))$ for all $a, v, \mathbf{y}$;

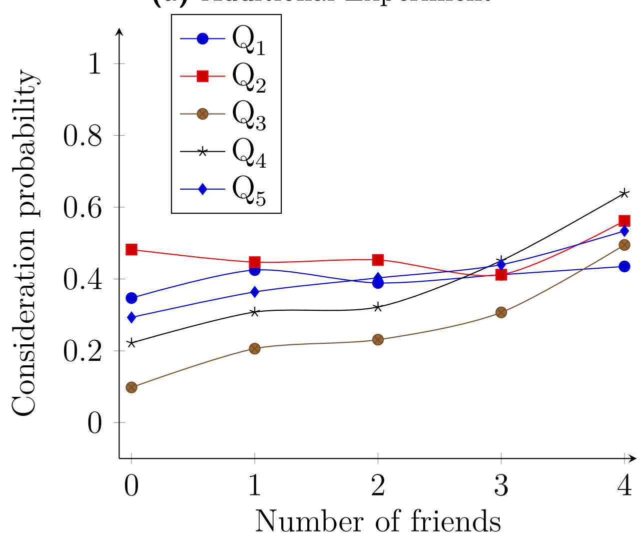
(ii) Model IIa: $\mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})) = \mathrm{Q}(\text{State}, \mathrm{N}_a^v(\mathbf{y}))$ for all $a, v, \mathbf{y}$. State equals to 1 if $v = y_a$, and 0 otherwise.

(iii) Model IIb: $\mathrm{Q}_a(v \mid y_a, \mathrm{N}_a^v(\mathbf{y})) = \mathrm{Q}_a(\text{State}, \mathrm{N}_a^v(\mathbf{y}))$ 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 9- 12 and Tables 15- 17.

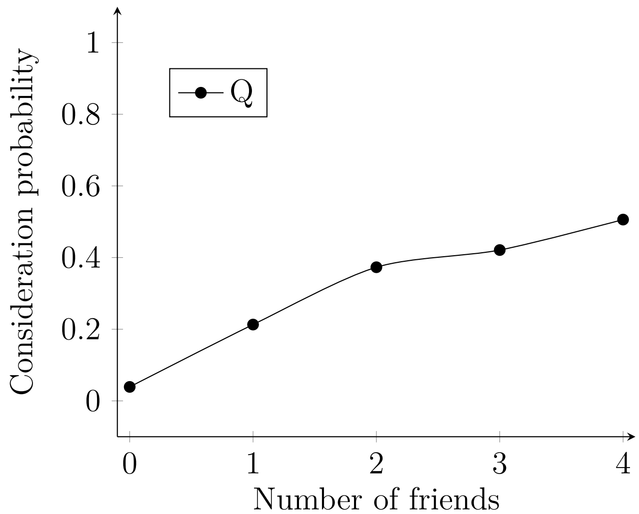


(a) Additional Experiment



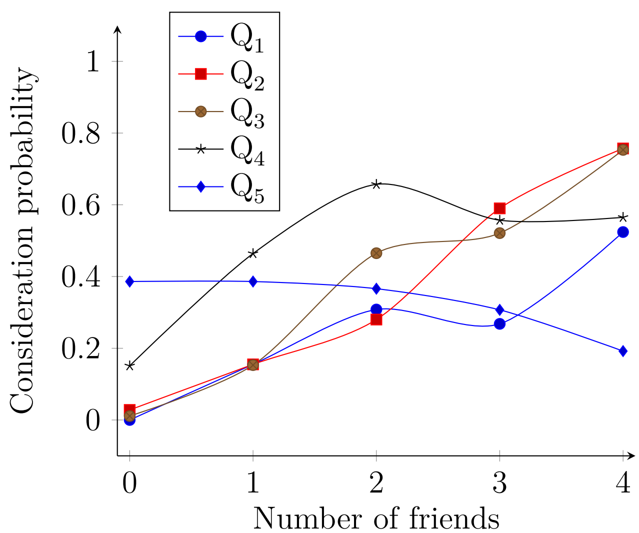
(b) Main Experiment

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



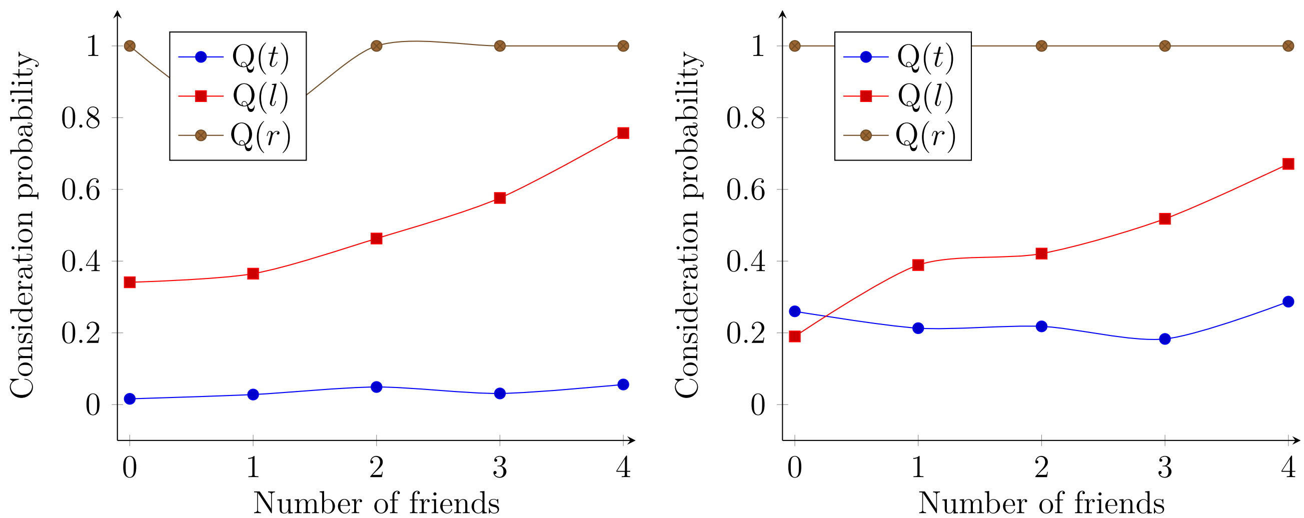
(a) Additional Experiment **(b)** Main Experiment

Figure 7 – Consideration probability. Model Ia. Second half of the samples.



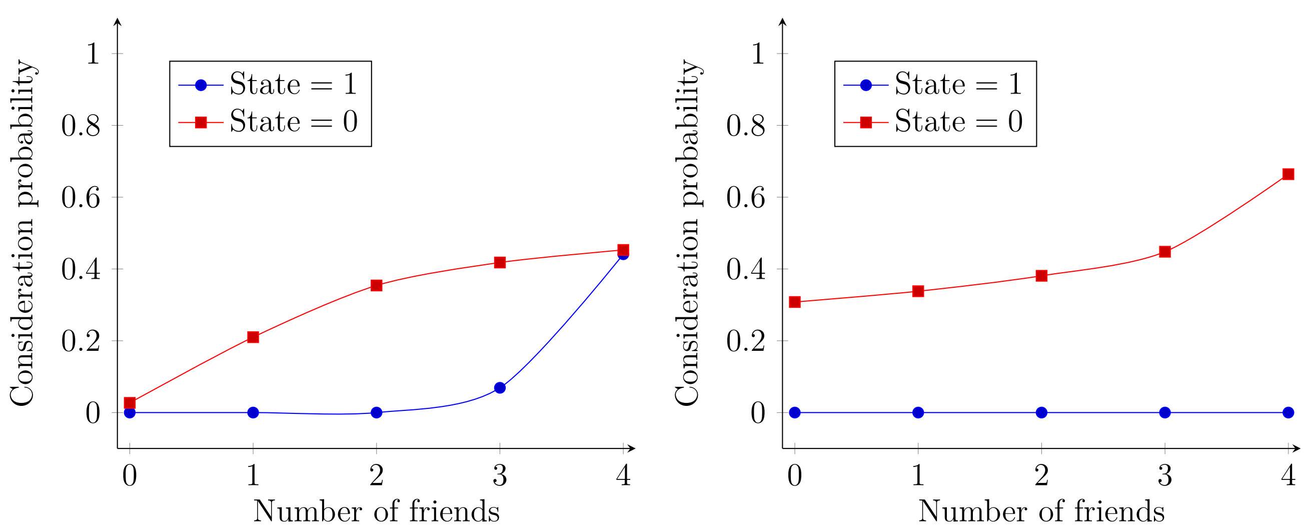
(a) Additional Experiment **(b)** Main Experiment

Figure 8 – Consideration probabilities. Model I(b). Second half of the samples.



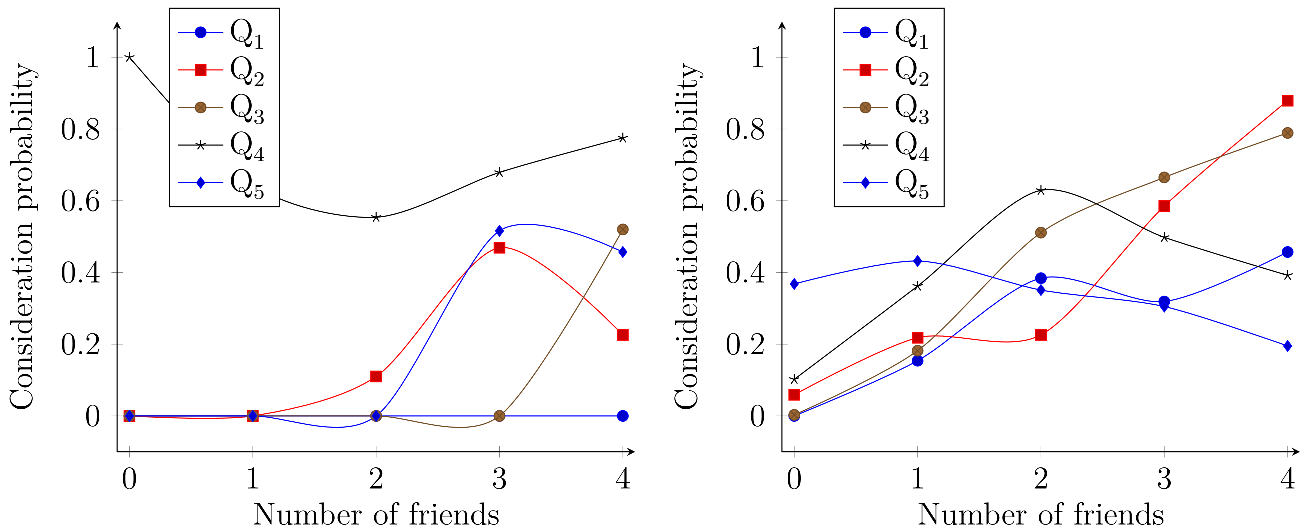
(a) Additional Experiment **(b)** Main Experiment

Figure 9 – Consideration probabilities. Model Ic.



(a) Additional Experiment **(b)** Main Experiment

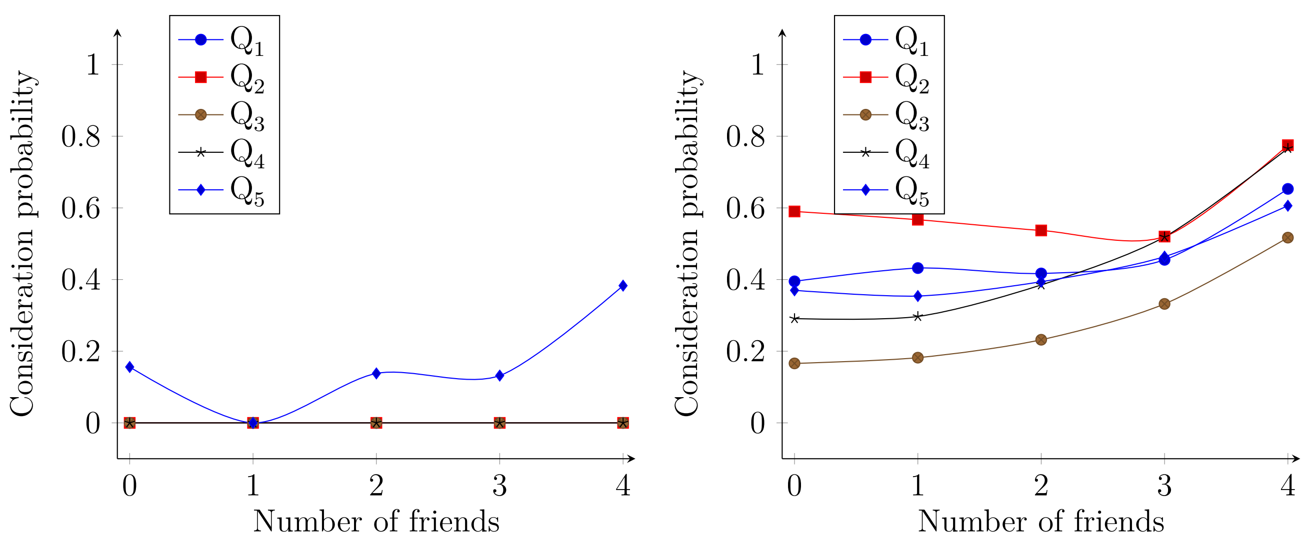
Figure 10 – Consideration probabilities. Model IIa.



(a) State= 1

(b) State= 0

Figure 11 – Consideration probabilities. Model IIb. Additional Experiment.



(a) State= 1

(b) State= 0

Figure 12 – Consideration probabilities. Model IIb. Main Experiment.

Table 14 – Preferences. Model I(b).

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

Table 15 – Preferences. Model Ic

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

Table 16 – Preferences. Model IIa

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

Table 17 – Preferences. Model IIb

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