Peer influence within clusters and stochastic choices: Mechanism, identification, and characterization

Abhinash Borah, Raghvi Garg, Ojasvi Khare and Siddarth Venkatesh*

December 19, 2024

Abstract

We introduce a stochastic choice model of peer influence in the context of a cluster network. Our goal is to propose a mechanism through which such peer influence works within clusters and show that the mechanism permits an exact behavioral identification of the clusters that form in society, along with individual preferences and idiosyncratic susceptibilities to influence. Under our mechanism that draws on the idea of influence-based random consideration, the more frequently an alternative is perceived to be chosen in a cluster, the greater the probability that any individual in that cluster considers it. An individual's choice probabilities, in turn, are determined from these consideration probabilities according to the random consideration set rule of Manzini and Mariotti (2014). Influence within clusters is mutual, and the choice procedure we introduce captures the steady-state choice profile of this interactive decision-making process. We establish that these interactions are mutually consistent and such profiles exist. Further, we behaviorally characterize the procedure.

JEL codes: D01, D91

Keywords: peer influence; clusters; influence-based random consideration; stochastic choices; interactive choice procedure and behavioral foundation; behavioral identification of networks

*Borah and Venkatesh: Ashoka University (e-mails: abhinash.borah@ashoka.edu.in, siddarth.venkatesh_tf@ashoka.edu.in); Garg: Indian Institute of Technology Kanpur (e-mail: raghvi@iitk.ac.in); Khare: Indian Statistical Institute, Delhi (e-mail:ojasvi19r@isid.ac.in). Correspondence address: Department of Economics, Ashoka University. Rajiv Gandhi Education City, Sonepat. Haryana - 131029, India.

1 Introduction

Understanding the influence of one's peers on behavior, including understanding the exact channel and mechanism through which such influence works, is an enduring question in economics. At the same time, if such peer effects exist, an analyst or outside observer may be greatly interested in the question of whether the social network underlying these peer effects can be uniquely identified from behavior. Moreover, since preferences and behavior may differ in the presence of influence, the analyst may also be interested in unearthing the preferences of different individuals, along with their idiosyncratic susceptibilities to influence. In this paper, we propose a stochastic choice model of peer influence that addresses these questions in the context of a cluster network, that is, a social network in which the set of individuals are partitioned into clusters, and all individuals within a cluster are connected.[1] Specifically, we propose a fairly natural mechanism through which peer effects work that draws on the idea of influence-based random consideration.[2] Such random consideration generates stochastic choices. We show how this stochasticity in behavior can be meaningfully exploited to uniquely identify the clusters along with individual preferences and susceptibility to influence. We also behaviorally characterize the *interactive* choice procedure resulting from this process of mutual influence.

Why clusters? There are at least a couple of reasons why we focus on a cluster network. The first is because such networks are empirically relevant and are seen to form in many important societal contexts, especially ones in which peer influence is salient. For instance, interactions on social media often result in the formation of clusters. Using social networks formed on various social media platforms, studies have established a substantial level of similarity among users who are close to each other in the social network, leading to more cohesive and predictable behaviors among group members (Aiello et al., 2012). One such example is the network structure on Twitter formed by retweets, which are generally seen as acts of endorsement and support of the original content, is characterized by two distinct clusters formed along partisan lines (Conover et al., 2011). Indeed, much of politics in our current polarized times is characterized by rigidly defined partisan clusters. Clusters have also been reported in friendship networks. Further, notions of identity, based on categories like race, religion, and caste, are a natural source of clusterings in society. Given the salience of such clusterings, it is only natural that individual behavior is psychologically

[1]Formally, the underlying graph representing the network is the disjoint union of complete graphs.

[2]The observation that, in any given choice problem, only a subset of the available alternatives may receive a decision maker's consideration (the "consideration set") has been well-documented in the literature, e.g.,Masatlioglu, Nakajima, and Ozbay (2012), Eliaz and Spiegler (2011), Manzini and Mariotti (2014), Brady and Rehbeck (2016), Lleras et al. (2017), Caplin, Dean, and Leahy (2018), and Dardanoni et al. (2020), amongst others. More recently, a strand of the literature has explored the possibility that social influence may play a role in determining a decision maker's consideration set in the sense that she may only consider alternatives that her peers choose or they recommend, e.g., Borah and Kops (2018) and Kashaev, Lazzati, and Xiao (2023).

influenced by dominant modes of behavior prevalent within one's cluster. For instance, consider the context of the partisan clustering in American society formed along the Democratic–Republican divide. It has been pointed out that people's political behavior, e.g., voting behavior, does not necessarily follow their preferences but instead is influenced by dominant views that prevail within their respective clusters (Cohen, 2003; Barber and Pope, 2019; Macy et al., 2019; Miller and Conover, 2015). Interestingly, such influence tends to carry over to even matters like lifestyle choices, consumption behavior, and questions of scientific policy (Kahan, 2010; DellaPosta, Shi, and Macy, 2015; Van Bavel et al., 2024; Dimant, 2024). Elsewhere, clusters seen in friendship networks result in strong peer effects affecting a whole range of behavior, including formative ones like smoking, prosocial attitudes, aggressive behavior (Ehlert et al., 2020; Ellis and Zarbatany, 2017; Lodder et al., 2016). Adolescents in clusters with smokers are more likely to adopt smoking habits, and also tend to choose friends with similar smoking habits (Ennett and Bauman, 2000; Seo and Huang, 2012; Hall and Valente, 2007). Norms and social pressure among clusters formed by adolescents, particularly girls, strongly influence body image concerns and dieting behavior (Paxton et al., 1999). Positive peer influence within clusters has also helped students boost their academic performance (Zimmerman, 2003). In short, the consistent findings across these studies highlights the impact that clusters have in shaping and reinforcing behaviors through social influence, peer pressure, and shared norms.

The second reason for our focus on clusters is the attention that such networks have received in the econometrics literature looking at the identification of peer effects. In his seminal paper, Manski (1993) discusses the reflection problem, highlighting the challenges in identifying endogenous peer effects when such peers are embedded within clusters. The interactive nature of decision making makes it challenging to discern whether the similarity in behavior among peers is due to the influence of peers' outcomes (endogenous peer effects), peers' characteristics (contextual effects), or due to shared external factors (correlated effects). Indeed, Manski (1993) establishes that even under the assumption that the cluster network is known to the econometrician, identifying endogenous peer effects may not be possible. To disentangle these effects, the literature (Bramoullé, Djebbari, and Fortin (2009), De Giorgi, Pellizzari, and Redaelli (2010), Lin (2010)) has had to move away from cluster networks and consider peers embedded within a general network structure. In such networks, identification is possible by exploiting a novel identification strategy involving non-transitivity of peers under which peers of peers are not peers. The problem that we are looking at is different from that in Manski (1993). Whereas Manski looked at identifying the endogenous peer effect coefficient in a regression equation, assuming the knowledge of the cluster network, we consider the problem of the analyst not knowing the clusters and trying to identify them from behavior. Even then, a natural question that emerges is whether the difficulty that cluster networks pose for identification in the former problem is also true for the latter. One of our key findings in this paper is that under an appropriate specification of the mechanism underlying influence, the answer is no—exact identification of the cluster network is indeed possible from behavior.

What, then, is the mechanism underlying peer influence that we envisage in this paper that successfully addresses the identification challenge? Indeed, it is this mechanism that we now describe that drives a connection between peer influence and stochastic choices. Individuals in our world have well-defined preferences and would otherwise choose deterministically. However, susceptibility to influence tends to draw their attention more toward those alternatives perceived as popular and more frequently chosen by their peers in their cluster. Such alternatives, therefore, end up having a higher probability of being considered by any individual within that cluster. This higher probability of being considered translates into a higher probability of being chosen. This link captures the very plausible phenomenon of individuals being psychologically drawn towards behavior they view as popular and widely prevalent within their cluster. For instance, if vaccine hesitancy is a more probable mode of behavior amongst cultural conservatives and is, therefore, more likely to be observed in this cluster, then this mode of behavior is more likely to be considered and chosen by any individual within that cluster. In turn, the higher the probabilities with which individuals within a cluster choose an alternative, the stronger the perception that it is a popular and more frequently chosen option. Thus, this process of mutual influence results in a situation of *interactive* decision-making.

We draw on the random consideration set rule of Manzini and Mariotti (2014) to capture these interactions formally. In that model, an individual decision maker (DM) has preferences over a given set of alternatives. However, her choice behavior is intermediated by the probabilities with which alternatives are considered. These consideration probabilities in their model are exogenously specified and independent of the menu where these alternatives occur. Given these consideration probabilities, the DM's probability of choosing an alternative in a menu is given by the conjoint probability of the event that this alternative receives consideration, but alternatives strictly preferred to it do not. We adopt this same choice rule for DMs in our model but with the innovation that the consideration probabilities are *endogenously* determined by the interactive process of peer influence within clusters as outlined above. Our choice procedure captures the steady state choice profile resulting from this process of mutual influence and interactive decision making. We refer to such a choice profile as a *clustered choice profile.*

Given that our choice procedure is interactive, the first key question that needs to be answered is whether the system of interactions outlined in it is mutually consistent. In other words, we have to answer the question about the existence of a clustered choice profile for any configuration of the parameters of our model, specifically, the societal clusters, individual preferences, and individual-level influence parameters capturing the extent of susceptibility to influence. Our first key result shows that such choice profiles are guaranteed to exist.

As noted above, a highlight of our model is that it can be uniquely identified from choice data, an observation that we formally establish in the second key result of the paper. This is an attractive feature of the model, given that full identification of behavioral choice models is often not possible. Further, identifying the structure of the social network of interactions from behavior is often a challenging question for social scientists. Therefore, it is noteworthy that the model permits a precise identification of the clusters that form in society from observable behavior. Along with this, individual preferences and the extent of their susceptibility to influence are also exactly identified. This, therefore, allows an analyst to identify the precise extent to which behavior is determined by preferences and the extent to which it is influenced. Moreover, she is able to distinguish between DMs who are influenced less and ones who are influenced more, including DMs who are fully immune from influence and behave rationally in the standard sense and DMs who are highly influenced, with behavior that is not in line with their preferences in the sense that they prefer alternative a over b, but may be more likely to choose b than a under the influence of others in their cluster.

On the point of identification, a comparison with the identification of the random consideration set rule model of Manzini and Mariotti (2014) is worth noting. That model, too, is uniquely identified. However, the authors show that this is true only when the consideration probabilities are menu-independent. If, instead, these are menu-dependent, then the identification of the model is completely lost, and any stochastic choice data set can be rationalized by it. In our model, consideration probabilities are menu-dependent. However, the endogenous structure on them introduced through the process of peer influence is such that full identification is still possible.

The third important result in our analysis is a behavioral characterization of clustered choice profiles. We use the structure of stochastic choices to guide this exercise. Specifically, we look at any individual who chooses stochastically. Given that any such DM is otherwise rational, having well-defined preferences over alternatives, stochasticity in choices speaks to the working of influence within clusters. Moreover, since such influence is an interactive process, it connects patterns of stochastic/non-stochastic behavior within any such cluster. This observation motivates the key revealed elicitation in the model—DM 1 is *revealed to be influenced* by DM 2 if she is seen choosing an alternative in a menu for sure *only if* DM 2 chooses it for sure as well. In other words, the constraint on the behavior of a DM who is influenced is that she is unable to choose deterministically based on preference maximization, independent of what others are choosing. Rather, a necessary condition for her to choose her preferred alternative for sure is her "influencers" doing so as well—else she chooses stochastically in line with influence-based random consideration. We use this critical insight to identify who influences whom and, thus, identify the clusters in society. This paves the way for the behavioral characterization of the model.

We add to a relatively recent literature that has studied peer influence within a behavioral choice-theoretic framework. There are different channels through which influence has been modeled in this literature. For instance, influence may directly alter the DM's perception of alternatives and thus her preferences over them; or it may modify the alternatives she pays attention to in a menu. We adopt the latter approach in this paper. Like in Kashaev, Lazzati, and Xiao (2023), Lazzati (2020), and Borah and Kops (2018), we assume that the choices of a DM's peers impact the alternatives she considers in a menu. In both Kashaev, Lazzati, and Xiao (2023) and Lazzati (2020), similar to our model, a DM's probability of considering an alternative depends on the number of her peers choosing it. However, the formal development of their theory substantially differs from ours as they focus on a dynamic setup. In Kashaev, Lazzati, and Xiao (2023), an individual chooses an alternative from her random consideration set at a given point in time and then revises her choice in the future, drawing on the choices made by her peers. This leads to a sequence of joint choices that evolve according to a continuous-time Markov process. They leverage this to uniquely identify the parameters of the model. On the other hand, Borah and Kops (2018) consider a deterministic two-stage choice procedure. In the first stage, a DM pays attention to those alternatives chosen by a sufficient number of individuals who influence her. In the second stage, she chooses the best among the considered alternatives according to her preferences. Unlike our model, influence in theirs is unidirectional, and they model the influence structure using a directed graph. Chambers, Cuhadaroglu, and Masatlioglu (2023) incorporate social influence into the classic Luce model. Unlike ours, influence in their model is captured through changes in DMs' preferences instead of their consideration sets. Chambers, Masatlioglu, and Turansick (2021) also study joint choices made by individuals and consider the possibility of correlation among choices across individuals. Furthermore, they differentiate between individual choices influencing each other and the possibility that these choices are consistent with an underlying unobserved signal. Cuhadaroglu (2017) introduces a two-stage choice procedure where a DM first shortlists based on her preferences that may be incomplete. In the second stage, she draws on her peers' preferences to compare the remaining alternatives. Fershtman and Segal (2018) model social influence by employing two sets of preferences – private and observable behavioral preferences. A social influence function determines the behavioral preferences of a DM as a function of her private preferences and the observable behavioral preferences of others. The focus of their paper is to highlight certain properties of these social influence functions and their implications for equilibrium behavior.

The rest of the paper is organized as follows. The next section lays out the setup of the model. Section 3 formally defines the clustered choice procedure. It also provides our existence result establishing that such profiles exist. Section 4 provides our identification result establishing that the model is uniquely identified. Section 5 provides a behavioral characterization of the procedure. Proofs of results appear in the Appendix.

2 Setup

Let $I = \{1, \dots, n\}$ be a set of individuals in society, with typical individuals denoted by i, j, etc. There is a partition $\{N^1, \dots, N^S\}$ of I into clusters. For any such clustering $N = \{N^1, \dots, N^S\}$, let $N(i)$ denote the element of the partition that individual $i \in I$ belongs to. Note that $j \in N(i)$ iff $i \in N(j)$. To keep interactions in the model interesting, we assume all clusters are non-singleton. Let X denote a finite set of alternatives, with typical elements denoted by a, b, etc. Each individual decision maker (DM), $i \in I$, has preferences over the alternatives in X. These preferences are captured by a strict preference ranking $\succ_i \subseteq X \times X$.[3] Further, let $\mathcal{X}$ denote the set of non-empty subsets of X, with typical elements A, B, etc., which we refer to as menus.

It would seem natural that if such clusters form, there must exist some minimal connection between these clusterings and the preferences of individuals forming them. We assume a fairly weak condition connecting clusters to preferences that is in the spirit of preference-based homophily. Specifically, between any two clusters, there must be some menu on which the preferences of everyone within one cluster regarding the best alternative agree, but this agreement is not shared within the other cluster. In other words, if two sets of individuals are part of two different clusters, there exists some minimal cluster-level preference disagreement between them. To state the condition formally, first, we define a menu as *top agreeable* for a cluster $N^s \in N$ if the $\succ_i$-best alternative in the menu (the "top") for all $i \in N^s$ is the same.

Condition 2.1 (Homophily)**.** *For all pairs of clusters N^s, N^t, there exists a menu that is top agreeable for one but not top agreeable with the same top for the other*

Clusterings which have this property will be referred to as *homophilous clusterings* and going forward we restrict attention to such clusterings.

In our framework, in addition to her preferences, the behavior of any DM may be influenced by that of others in her cluster. This produces a situation of *interactive decision-making*. We model this interactive behavior using random choice rules. We allow for the possibility that a DM may not pick any alternative from a menu, so we also assume the existence of a default alternative a^*.[4] Denote $X^* = X \cup \{a^*\}$ and $A^* = A \cup \{a^*\}$ for all $A \in \mathcal{X}$.

Definition 2.1. A *random choice rule* is a map $p : X^* \times \mathcal{X} \to [0, 1]$ such that $\sum_{a \in A^*} p(a, A) = 1$ for all $A \in \mathcal{X}$, and $p(a, A) = 0$ for all $a \notin A^*$. The random choice rule is non-trivial if $\exists A \in \mathcal{X}$ s.t. $p(a, A) \neq 1$ for any $a \in A^*$.

The interpretation is that $p(a, A)$ is the probability that the alternative $a \in A^*$ is chosen when the DM faces the menu A. Further, if there exists some menu A in which the DM

[3] By a strict preference ranking, we mean a total, asymmetric and transitive binary relation.

[4] This modeling choice is the same as that made in Manzini and Mariotti (2014).

doesn't choose deterministically, i.e., $p(a, A) \neq 1$ for any $a \in A^*$, then we say that the random choice rule capturing her behavior is non-trivial(ly stochastic). In our set-up, the behavior of each individual $i \in I$ will be captured by a non-trivial random choice rule $p_i : X^* \times \mathcal{X} \to [0, 1]$. Finally, note that we will often abuse notation by suppressing set delimiters. For instance, we write a instead of $\{a\}$, ab instead of $\{a, b\}$, etc.

3 Choices within Clusters

We now formally describe the process of interactive decision-making brought about by the working of influence and how the profile of random choice rules within any cluster are *co-determined*. For this, we draw on the random consideration set model of Manzini and Mariotti (2014). The key feature of their model is that, in any menu, alternatives are not guaranteed to receive full attention or consideration. Instead, such consideration is stochastic, and every alternative is considered with some probability that may be less than 1. In their model, these probabilities are exogenous and menu-independent. We model peer influence through consideration probabilities but with the innovation that these are *endogenous*. Specifically, a DM's probability of considering an alternative in a menu depends on how likely are individuals in her cluster to choose it, thus introducing an endogeneity. This captures the intuitive idea that the more popular and frequent a choice is likely to be in the DM's cluster, the more attention it draws and more likely it is to receive consideration. This also means that consideration probabilities in our setting are *menu-dependent*.

Formally, the probability with which any individual i considers an alternative a in a menu A is influenced by how likely, on average, she believes this alternative to be chosen within her cluster, captured by the average probability, $\frac{1}{|N(i)|} \sum\limits_{j \in N(i)} p_j(a, A)$.[5] Think of this average as a summary statistic capturing i's perception of the likely popularity of alternative a in menu A in her cluster. However, not all individuals are susceptible to influence to the same extent. Let $\beta_i \in (0, 1)$ be the probability of individual i's consideration in any menu being immune from influence so that $1 - \beta_i$ captures the probability of her being susceptible to influence. Accordingly, the probability of i considering any alternative in a menu is determined as follows: with probability β_i, the alternative receives full consideration, and with complementary probability $1 - \beta_i$, the consideration is equal to the average probability of this alternative being chosen in her cluster. That is, the consideration probability $\gamma_i(a, A)$ of i considering an alternative a in menu A is given by:

$$\gamma_i(a, A) = \beta_i + (1 - \beta_i) \frac{\sum\limits_{j \in N(i)} p_j(a, A)}{|N(i)|}$$

[5] $|N(i)|$ denotes the cardinality of $N(i)$. The choice probabilities $p_j(a, A), j \in N(i)$, are, of course, endogenous.

The probability with which any DM considers an alternative in a menu depends on the average probability of this alternative being chosen in her cluster. At the same time, what these choice probabilities are would depend on consideration probabilities. In other words, choice probabilities both determine consideration probabilities and, in turn, are determined by them, along with preferences in an environment of mutual influence and interactions within clusters. Our choice procedure captures the steady state of these interactions.

Definition 3.1. *The profile of non-trivial random choice rules $(p_i)_{i\in I}$ is a clustered choice if there exists*

(i) a collection of strict preference rankings $\{\succ_i : i \in I\}$,

(ii) a homophilous clustering $N = \{N^1, \ldots, N^S\}$ of I, $|N^s| > 1$, $\forall\ N^s \in N$, and

(iii) immunity from influence coefficients $\{\beta_i \in (0,1) : i \in I\}$

such that for any $i \in I$, $A \in \mathcal{X}$ and $a \in A$:

$$p_i(a, A) = \gamma_i(a, A) \prod_{b\in A: b\succ_i a} \big(1 - \gamma_i(b, A)\big) \tag{1}$$

and

$$\gamma_i(a, A) = \beta_i + (1 - \beta_i)\frac{\sum\limits_{j\in N(i)} p_j(a, A)}{|N(i)|} \tag{2}$$

The notion of a clustered choice profile can be thought of as a steady state of the process of mutual influence that takes place within clusters. Equation 1 specifies the random consideration set rule of Manzini and Mariotti (2014). Under it, for any $i \in I$, and any menu A, the probability of choosing the most preferred alternative according to $\succ_i$ in the menu, say a_1, is simply the probability that it is considered, $\gamma_i(a_1, A)$; the second most preferred alternative, say a_2, is the product of the probability that it is considered, but a_1 is not, $\gamma_i(a_2, A)(1 - \gamma_i(a_1, A))$; the third most preferred alternative, say a_3, is the product of the probability that it is considered, but a_1 and a_2 are not, $\gamma_i(a_3, A)(1 - \gamma_i(a_1, A))(1 - \gamma_i(a_2, A))$; and so on and so forth. The innovation of our model lies in how peer influence within clusters determines the consideration probabilities specified in equation 2, with these probabilities being endogenous and part of a process of mutual interactions as discussed above. Observe that, like in any steady state or equilibrium notion, the definition of a clustered choice profile requires DMs to hold correct beliefs, specifically about the average probability of different alternatives being chosen in any menu in their cluster.

3.1 Existence and Stability

Given that our choice procedure is one of interactive decision making, we need to ascertain that the interactions are mutually consistent. The following result establishes this by showing that a unique clustered choice profile exists for any specification of the parameters of the model.

Theorem 3.1. *A clustered choice profile exists for any collection of strict preference rankings $\{\succ_i : i \in I\}$, homophilous clustering $N = \{N^1, \ldots, N^S\}$ of I, and immunity from influence coefficients $\{\beta_i : i \in I\}$. Further, this profile is unique.*[6]

Proof: Please refer to Section A.1.

The result is established by an application of the Banach fixed-point theorem. Naturally, then, we can think about clustered choice profiles as the limit of an updating process of random choice rules. To do so, suppose that agents choose according to an initial random choice rule $p^0 = \left(p_i^0\right)_{i\in I}$. Let these agents have preferences $\{\succ_i : i \in I\}$, immunity from influence coefficients $\{\beta_i : i \in I\}$, and be partitioned into a homophilous clustering $N = \left\{N^1, \ldots, N^S\right\}$. Suppose that agents update consideration probabilities of alternatives in every successive period using a rule of thumb. Particularly,

$$\gamma_i^{t+1}(a,\, A) = \beta_i + (1-\beta_i)\frac{\sum_{j\in N(i)} p_j^t(a,\, A)}{|N(i)|}$$

for $A \in \mathcal{X}$, $a \in A$. Their probability of choosing said alternative in period $t+1$ is determined by

$$p_i^{t+1}(a,\, A) = \gamma_i^{t+1}(a,\, A) \prod_{a'\in A: a' \succ_i a} (1-\gamma_i^{t+1}(a',\, A))$$

Finally, the residual probability is assigned to the default alternative,

$$p_i^{t+1}(a^*,\, A) = 1 - \sum_{a\in A} p_i^{t+1}(a,\, A) = \prod_{a\in A}(1-\gamma_i^{t+1}(a,\, A))$$

From the proof of Theorem 3.1, we can establish that the sequence of choice rules obtained via this updating process converge to a unique steady state, the clustered choice profile associated with these parameters, irrespective of initial rule p^0.

Corollary 3.1. *Let $p^0 = \left(p_i^0\right)_{i\in I}$ be some collection of random choice rules. Let $\{\succ_i : i \in I\}$ be a collection of preferences, $N = \left\{N^1, \ldots, N^S\right\}$ a homophilous clustering, and $\{\beta_i : i \in I\}$ a collection of immunity from influence coefficients that represent the clustered choice profile $p^* = (p_i^*)_{i\in I}$. Define for each $i \in I$, γ_i^{t+1} as*

$$\gamma_i^{t+1}(a,\, A) \equiv \beta_i + (1-\beta_i) \sum_{j\in N(i)} \frac{\sum_{j\in N(i)} p_j^t(a,\, A)}{|N(i)|}$$

[6] It should be pointed out that the existence result does not rely on the clustering satisfying the weak homophily condition. The proof goes through even if this condition doesn't hold.

and correspondingly define p_i^{t+1} *as*

$$p_i^{t+1}(a,\, A) \equiv \gamma_i^{t+1}(a,\, A) \prod_{a' \in A : a' \succ_i a} (1 - \gamma_i^{t+1}(a',\, A))$$

for $a \in A$*, and* $p_i^{t+1}(a^*,\, A) \equiv 1 - \sum_{a \in A} p_i^{t+1}(a,\, A)$*, for each* $A \in \mathcal{X}$*. Then* $\lim_{t\to\infty} p^t = p^*$

4 Identification

We now address the important question about the identification of the clustered choice model. A key highlight of the model is that it can be uniquely identified from choice data. Identifying the structure of social interactions from observed behavior remains a challenging question for analysts. The fact that, in the clusterted choice model, the clusters that form, along with the preferences of the individuals in society and their respective immunity/susceptibility to influence, can be uniquely identified is a noteworthy property of the model.

Theorem 4.1. *Let* $(p_i)_{i\in I}$ *be a clustered choice profile. If* $\left\langle \{\succ_i : i \in I\}, N = \{N^1, \ldots, N^S\}, \{\beta_i : i \in I\} \right\rangle$ *and* $\left\langle \{\hat{\succ}_i : i \in I\}, \hat{N} = \{\hat{N}^1, \ldots, \hat{N}^T\}, \{\hat{\beta}_i : i \in I\} \right\rangle$ *are both clustered choice representations of* $(p_i)_{i\in I}$*, then* $N = \hat{N}$*, and for all* $i \in I$*,* $\succ_i = \hat{\succ}_i$*,* $\beta_i = \hat{\beta}_i$*.*

Proof: Please refer to Section A.3.

The key insight underlying the result is the observation that the identification problem essentially reduces to the analyst's ability to correctly identify the clusters. Once the clusters are identified, exact identification of the preferences and influence parameters is straightforward. Specifically, suppose the analyst has succeeded in identifying the clustering as $(N_i)_{i\in I}$. She can then determine for any $i \in I$ and any menu A, the average probability with which any alternative $a \in A$ is chosen in i's cluster. Denote this by $\hat{\mu}_i(a, A) = \frac{1}{N(i)} p_i(a, A)$. Clearly, as long as the clusters are uniquely identified, these average choice probabilities are as well. In the clustered choice model, information about these average choice probabilities are all that is needed by the analyst to back out the preferences of the individuals and their immunity from influence coefficients. As we formally establish in Lemma A.2 and Corollary A.3 and A.4, for any $i \in I$ and $a, b \in X$,

$$a \succ_i b \iff \frac{1 - p_i(a, ab)}{1 - p_i(b, ab)} \leq \frac{1 - \hat{\mu}_i(a, ab)}{1 - \hat{\mu}_i(b, ab)}$$

In other words, the notion of revealed preference in the model is the following. For any individual i, and any pair of alternatives, $a, b \in X$, if the ratio of the probability of i not choosing a to not choosing b is no greater than the corresponding ratio for her cluster, then

i prefers a to b. Accordingly, as soon an the analyst is able to identify any individual's cluster, she is able to identify her preferences as well.

A similar observation applies to any individual i's immunity from influence coefficient β_i. Consider any menu A in which i chooses stochastically, and let $a = \max_{\succ_i} A$. Then, it is straightforward to derive that i's susceptibility to influence, $1 - \beta_i$ is given by,

$$1 - \beta_i = \frac{1 - p_i(a, A)}{1 - \hat{\mu}_i(a, A)}$$

The numerator, $1 - p_i(a, A)$ gives the probability of i not choosing a in A despite it being her most preferred alternative. The denominator specifies the average probability of a not being chosen in A in her cluster. Therefore, the ratio gives a measure of the extent to which a not being chosen in her cluster translates to i not choosing it, despite it being her most preferred alternative in A. Therefore, it reveals the extent of her susceptibility to influence. Accordingly, as long as an individual's cluster and her preferences can be uniquely identified by the analyst, she can also uniquely identify her immunity from influence coeffieceint β_i

Therefore, the problem of uniquely identifying the parameters of the model boils down to the analyst's ability to identify the clusters. What makes such identification possible? As the reader may have guessed, a key feature that makes such identification possible is the presence of preference homophily in the clustering. Despite the weak form in which the model assumes homophily, in the presence of random consideration, it is sufficient to provide the analyst enough information to uniquely identify the clusters. To see what may go wrong w.r.t. identification when the clusterings are not homophilous, consider the following example.

Example 4.1. Suppose $X = \{a, b, c\}$ is the set of alternatives and $I = \{1, 2, 3, 4\}$ the set of individuals in society who are partitioned into the clustering $N = \{\{1, 3\}, \{2, 4\}\}$. Further let their preferences be given by: $a \succ_1 b \succ_1 c$, $a \succ_2 b \succ_2 c$, $a \succ_3 c \succ_3 b$, $a \succ_4 c \succ_4 b$; and their immunity from influence coefficients by $\beta_1 = \beta_2 = 3/5$ and $\beta_3 = \beta_4 = 2/5$. Observe that the clustering is not homophilous, as there is no menu that is top agreeable for one of the clusters but not for the other with the same top.[7] Applying the structure of interactions underlying a clustered choice profile (equations 1 and 2 of Definition 3.1) gives us the following steady state profile: $p_i(a, A) = 1$ for all $i \in I$ and A such that $a \in A$, and for the menu $\{b, c\}$, choice probabilities are given by,

	p_1	p_2	p_3	p_4
b	$\frac{\sqrt{1801}-31}{14}$	$\frac{\sqrt{1801}-31}{14}$	$\frac{3\sqrt{1801}-107}{28}$	$\frac{3\sqrt{1801}-107}{28}$
c	$\frac{\sqrt{1801}-4}{51}$	$\frac{\sqrt{1801}-4}{51}$	$\frac{\sqrt{1801}-21}{34}$	$\frac{\sqrt{1801}-21}{34}$

[7] Any menu A with $a \in A$ is top-agreeable with a as the top for both clusters. The remaining non-singleton menu $\{b, c\}$ is not top-agreeable for either of the clusters.

However, since individuals 1 and 2 are replicas of one another, as are 3 and 4, the clustering given by $\hat{N} = \{\{1, 4\}, \{2, 3\}\}$, along with the same preferences and influence coefficients, also represents the same profile of non-trivial random choice rules as a steady state of clustered choice interactions.

What is it that frustrates the analyst's attempt to exactly identify the clusters in this example? To understand this, we need to understand that what permits exact identification is a particular pattern of stochastic and deterministic behavior amongst individuals in society. Since all individuals have a well-defined preference ranking over the set of alternatives, they would in the absence of influence deterministically choose the best alternative in any menu according to their preferences. Hence, the presence of stochastic behavior suggests the possibility of influence. At the same time, though, it is not as if individuals always choose stochastically. However, for them to choose an alternative with probability one from a menu under a clustered choice profile, everyone else in their cluster must do so as well (Corollary A.1). In other words, the analyst may get an "upper bound" of an individual's cluster by identifying all those who choose an alternative from a menu with probability one whenever she does so. But to get exact identification, she needs to be assured that there is at least one menu where only those in her cluster choose an alternative deterministically along with her and no one outside her cluster does so. It is precisely here that homophilous clusterings comes to the analyst's aid in identification. As we formally establish in Lemma A.1 and Corollary A.1, a menu is top agreeable for a cluster *if and only if* everyone in the cluster chooses the top with probability one. Homophilous clusterings, by ensuring that between any two clusters there exists at least some menu that is top agreeable for one of the clusters but not for the other with the same top, guarantees that exact identification of clusters is possible. Going back to the example, we see this clearly as the menus in which individuals choose deterministically are the same for all of them. These are precisely those menus that include alternative a, which is chosen with probability one in all of them. Based on this choice data, the best that the analyst can do is to over-identify everyone's cluster as all of society. However, if the underlying clustering is not homophilous, she can't improve upon this upper bound any further.

Nevertheless, it is also true that not imposing weak homophily does not necessarily inhibit unique identification of clusters. This is seen in the following example.

Example 4.2. Suppose $X = \{a, b, c\}$ is the set of alternatives and $I = \{1, 2, 3, 4\}$ the set of individuals in society. Let the profile of non-trivial random choice rules for these four individuals be given by $p_i(a, A) = 1$ for $A = \{a,b\}, \{a,c\}, \{a,b,c\}$, $i \in I$, and for the menu $\{b, c\}$ choice probabilities given by the following:

	p_1	p_2	p_3	p_4
b	$1/2$	$3/5$	$2/5$	$1/2$
c	$\frac{31-\sqrt{805}}{26} \approx 0.101$	$\frac{150-4\sqrt{805}}{481} \approx 0.076$	$\frac{78-3\sqrt{341}}{67} \approx 0.337$	$\frac{457-15\sqrt{341}}{1139} \approx 0.158$

If this choice data has to be consistent with the interactions underlying clustered choices, then note first that for any menu where an individual chooses stochastically,

$$\gamma_i(a,\, A) = \beta_i + (1-\beta_i)\hat{\mu}_i(a,\, A) > \hat{\mu}_i(a,\, A),$$

and their best alternative in the menu is chosen with the consideration probability. This immediately tells us that $\{1,4\}$ cannot form a cluster because there is no such alternative for 1 in the menu $\{b,c\}$, where her choice probability of choosing that alternative is greater than the average choice probability of 1 and 4 of choosing that alternative. By the same argument, it also follows that {1,2,3,4} cannot be part of a cluster. Note that neither $p_1(b,\, bc)$ nor $p_1(c,\, bc)$ would be greater than the averages across the four individuals.

Now look at the possibility of $\{1,3\}$ forming a cluster. Since $p_1(b,\, bc) > p_3(b,\, bc)$ and $p_3(c,\, bc) > p_1(c,\, bc)$, we can conclude that $b \succ_1 c$ and $c \succ_3 b$ if 1 and 3 are in the same cluster. Since $p_3(b,\, bc) = 2/5$ and $p_1(b,\, bc) = 1/2$, the values for $p_1(c,\, bc)$ and $p_3(c,\, bc)$ must satisfy the following equations. The first of them is

$$\frac{p_3(c,\, bc) - \frac{1}{2}\left(p_1(c,\, bc) + p_3(c,\, bc)\right)}{1 - \frac{1}{2}(p_1(c,\, bc) + p_3(c,\, bc))} = \frac{\frac{2}{5(1-p_3(c,bc))} - \frac{9}{10}}{\frac{11}{20}}$$

The second is

$$\frac{2p_1(c,\, bc) - \frac{1}{2}\left(p_1(c,\, bc) + p_3(c,\, bc)\right)}{1 - \frac{1}{2}\left(p_1(c,\, bc) + p_3(c,\, bc)\right)} = \frac{\frac{1}{2} - \frac{9}{20}}{\frac{11}{20}} = \frac{1}{11}$$

The above equations come from rearranging the consideration probability equations for each of the alternatives to yield an expression for β_i in terms of a menu A and some alternative $x \in A$

$$\beta_i = \frac{\gamma_i(x,\, A) - \hat{\mu}_i(x,\, A)}{1 - \hat{\mu}_i(x,\, A)}$$

Since this is equal for any $x,\, y \in A$, the above equations are obtained by equating two such expressions for β_i. It can then be verified that there is only one solution for this system of equations such that $(p_1(c,\, bc), p_3(c,\, bc)) \in [0,\, 1]^2$, which requires $p_1(c,\, bc) = \frac{457-15\sqrt{341}}{1139}$. However, this does not hold true for this profile, implying that 1 and 3 cannot be part of the same cluster either. This leaves $N = \{\{1,\, 2\}\,,\, \{3,\, 4\}\}$ as the only possible clustering.

Consider preferences
1: $a \succ_1 c \succ_1 b$
2: $a \succ_2 b \succ_2 c$

3: $a \succ_3 c \succ_3 b$
4: $a \succ_4 b \succ_4 c$
Then note that for values of $\beta_1 = \frac{847-29\sqrt{805}}{477+45\sqrt{805}} \approx 0.014$, $\beta_2 = 1/9$, $\beta_3 = \frac{869-36\sqrt{341}}{495+66\sqrt{341}} \approx 0.119$, $\beta_4 = 1/11$, it can be verified that $p_i(x, A) = \gamma_i(x, A) \prod_{y \succ_i x}(1-\gamma_i(y, A))$ with $\gamma_i(x, A) = \beta_i + (1-\beta_i)\hat{\mu}_i(x, A)$. Then, this profile has a clustered choice representation, and by the uniqueness of the clusters, is unique.

However, note that this representation does not satisfy the weak homophily condition, as the set of top-agreeable menus for both clusters is $\{\{a, b\}, \{a, c\}, \{a, b, c\}\}$, with the tops being the same. This shows that weak homophily is not a necessary restriction on the representation for the unique identification of parameters.

Finally, we observe that an analyst need not obtain choice data on all menus to be able to identify the parameters that generate a clustered choice profile. Particularly, we show that it is enough if the analyst has choice data on $\bar{\mathcal{X}} = \{A \in \mathcal{X} : |A| = 2\}$, the set of all binary menus. If the data originates from our model, the underlying parameters can be identified uniquely.

Proposition 4.1. *Suppose $(p_i)_{i\in I}$ is a clustered choice profile and data on $(p_i)_{i\in I}$ is available for $\bar{\mathcal{X}}$. Then parameters that represent $(p_i)_{i\in I}$ can be identified uniquely.*

Proof: Please refer to Appendix Section A.4

5 Characterization

In our set-up, each individual has a single set of well-defined preferences; hence, without any other consideration or influence, she should choose non-stochastically. Conversely, her choosing stochastically reveals that others may influence her behavior, given the random consideration that such influence generates. That is, the constraint imposed by influence on behavior is that it prevents individuals from *independently* choosing an alternative for sure from a menu like a standard DM. This reasoning suggests a simple way to behaviorally determine who is connected to whom in the influence network. Since the manifestation of influence is that it introduces stochasticity into choices, a simple way to determine that an individual j influences another individual i is if we see that i chooses some alternative a in a menu for sure *only if* j does so. This is because if some such j who influences i happens to choose a with a probability of less than one, she has a positive probability of choosing some other alternative b in the menu. But by doing so, she would draw i's attention towards b, and introduce a *chance* of i choosing this alternative, thus ensuring that a is not chosen for sure. Of course, the argument is symmetric. i not choosing a for sure would influence j not doing so either. Formally, any $i, j \in I$ are (revealed to be) *connected* if for any $A \in \mathcal{X}$ and $a \in A$, $p_i(a,\, A) = 1$ if and only if $p_j(a,\, A) = 1$. Our first axiom makes the point that no individual is an island and everyone is connected.

Axiom 1 (**Connectedness**). *For each $i \in I$, there exists $j \neq i$ such that i and j are connected.*

Denote the set of individuals to whom $i \in I$ is connected by $R(i)$. That is,

$$R(i) = \{j \in I : p_i(a,\, A) = 1 \iff p_j(a,\, A) = 1, A \in \mathcal{X},\, a \in A\}$$

Further, define for any $A \in \mathcal{X}$ and $a \in A$, the average choice probability of a being chosen in A in $R(i)$ by:

$$\mu_i(a, A) = \frac{1}{|R(i)|} \sum_{j \in R(i)} p_j(a, A)$$

Our next axiom delves into the question of when does a DM's behavior reveal that she has consistent preferences. Suppose, for some menu A and $a \in A$, $p_i(a, A) \geq p_i(b, A)$ or, equivalently, $\frac{1-p_i(a,A)}{1-p_i(b,A)} \leq 1$, for all $b \in B \setminus a$. From this, can we infer that a is i's most preferred alternative in A? Well, not necessarily, because the choice probabilities, $p_i(a, A)$ and $p_i(b, A)$, may reflect not just i's preference but also the relative popularity of a vs. b amongst i's connections, as captured by $\mu_i(a, A)$ and $\mu_i(b, A)$. Hence, to make a correct inference, one needs to take cognizance of this. One may do so by comparing the probability ratio of i not choosing a to not choosing b, $\frac{1-p_i(a,A)}{1-p_i(b,A)}$, to the corresponding ratio capturing average choice behavior amongst her influencers, $\frac{1-\mu_i(a,A)}{1-\mu_i(b,A)}$. In particular, if the first ratio is no larger than the second for any $b \in A \setminus a$, then it reveals that a is indeed i's most preferred alternative in A. Our next axiom draws on the standard independence of irrelevant alternatives (IIA) condition for deterministic choices to introduce the requirement that such inferences about a DM's preferences should not be menu dependent.

Axiom 2 (**Stochastic IIA**). *For all $i \in I$ and $A \in \mathcal{X}$, $\exists!$ $a \in A$ such that $\frac{1-p_i(a,A)}{1-p_i(b,A)} \leq \frac{1-\mu_i(a,A)}{1-\mu_i(b,A)}, \forall b \in A \setminus a$; and for any $B \subseteq A$, with $a \in B$, $\frac{1-p_i(a,B)}{1-p_i(b,B)} \leq \frac{1-\mu_i(a,B)}{1-\mu_i(b,B)}, \forall b \in B \setminus a$.*

That is, in any menu A, there exists an alternative $a \in A$ that is revealed to be the most preferred, and in any sub-menu B of A in which a is present, the same is true.

Our final axiom seeks to elicit from behavior the extent to which a DM is susceptible to influence; and it imposes an independence requirement on such idiosyncratic susceptibility to influence. To understand the idea behind the axiom, note that for any menu A in which i chooses stochastically, if $a_1 \in A$ is i's most preferred alternative in the menu, then the ratio $\frac{1-p_i(a_1,A)}{1-\mu_i(a_1,A)}$ captures the extent of i's susceptibility to influence when it comes to her likelihood of choosing a_1. For instance, suppose $p_i(a_1, A) = 0.9$ and $\mu_i(a_1, A) = 0.6$. Then it means that a 40% chance of a_1 not being chosen on average amongst her connections translates into a 10% chance of this alternative not being chosen by i, even though it is her most preferred alternative. Hence, the ratio $\frac{0.1}{0.4} = \frac{1}{4}$ captures i's idiosyncratic susceptibility to influence. Now consider i's second most preferred alternative in A, call it a_2. Note that w.r.t. a_2, $\frac{1-p_i(a_2,A)}{1-\mu_i(a_2,A)}$ is not the correct measure of idiosyncratic susceptibility to influence. This is because a_2 is not i's most preferred alternative in A and that is part of the reason why this alternative may not be chosen. Therefore, the correct probability to look at is not the unconditional probability of a_2 not being chosen, $1 - p_i(a_2, A)$, but rather the probability that a_2 is not chosen conditional on no alternative preferred to it, i.e., a_1, being chosen; specifically, the probability $1 - \frac{p_i(a_2,A)}{1-p_i(a_1,A)}$. The susceptibility to influence w.r.t. the choice of a_2 is then captured by the ratio $\frac{1-\frac{p_i(a_2,A)}{1-p_i(a_1,A)}}{1-\mu_i(a_2,A)}$. If this susceptibility to influence is independent of the alternatives under consideration, then it should be the case that $\frac{1-p_i(a_1,A)}{1-\mu_i(a_1,A)} = \frac{1-\frac{p_i(a_2,A)}{1-p_i(a_1,A)}}{1-\mu_i(a_2,A)}$. Our final axiom requires that this idiosyncratic susceptibility to influence should not vary across alternatives in a menu, nor should it vary across menus. To capture this formally, we need to introduce some notation. First, for any menu A in which i chooses stochastically, i.e., $p_i(b, A) \neq 1$, for any $b \in A^*$, and $a \in A$, define:

$$\bar{A}_i(a) = \left\{ b \in A : \frac{1 - p_i(b, ab)}{1 - p_i(a, ab)} \leq \frac{1 - \mu_i(b, ab)}{1 - \mu_i(a, ab)} \right\}$$

Based on our discussion above, $\bar{A}_i(a)$ contains those alternatives in A that are revealed to be preferred by i to a. Then the probability of a being chosen conditional on no alternative that is revealed to be preferred to it being chosen is given by:

$$p_i^*(a, A) = \frac{p_i(a, A)}{1 - \sum_{b \in \bar{A}_i(a)} p_i(b, A)}$$

We can now formally state the axiom.

Axiom 3 (**Menu independence of influence**). *For any $i \in I$, menus $A, B \in \mathcal{X}$ in which i chooses stochastically, and $a \in A$, $b \in B$,*

$$\frac{1 - p_i^*(a, A)}{1 - \mu_i(a, A)} = \frac{1 - p_i^*(b, B)}{1 - \mu_i(b, B)} < 1$$

These three axioms together constitute a behavioral foundation for the clustered choice model.

Theorem 5.1. *A profile of random choice rules $(p_i)_{i \in I}$ is a clustered choice if and only if it satisfies connectedness, stochastic IIA, and menu independence of influence.*

Proof: Please refer to Section A.5.

6 Comments

6.1 Comparison with Manzini and Mariotti (2014)

Since we draw closely on Manzini and Mariotti (2014), an important observation with respect to the consideration probabilities in the two models is in order. The consideration probabilities in Manzini and Mariotti (2014) are exogenous and menu-independent. For this set-up, they characterize the behavioral restrictions implied by their random consideration set rule model and show that the model is uniquely identified. On the other hand, our model endogenizes consideration probabilities and these are *menu-dependent*. Manzini and Mariotti (2014) show that for menu-dependent consideration probabilities, their model becomes too permissive. Formally, a menu-dependent random consideration set rule is a random choice rule $p_{\succ,\delta}$ for which there exists a pair $(\succ, \delta)$, where $\succ$ is a strict preference ranking on X and $\delta : X \times \mathcal{X} \to (0,1)$ is a menu-dependent consideration probabilities mapping, s.t.,

$$p_{\succ,\delta}(a, A) = \delta(a, A) \prod_{b \in A: b \succ a} (1 - \delta(b, A)), \text{ for all } A \in \mathcal{X}, \text{ for all } a \in A$$

Theorem (Manzini and Mariotti (2014))**.** *For every strict preference ranking $\succ$ on X and for every random choice rule p, there exists a menu-dependent random consideration rule $p_{\succ,\delta}$ such that $p = p_{\succ,\delta}$.*[8]

In other words, the model imposes no restriction on choice data. Because of this, Manzini and Mariotti (2014) write: "So, once we allow the attention parameters to be menu-dependent, not only does the model fail to place any observable restriction on choice data, but the preference relation is also entirely unidentified. Strong assumptions on the function δ are needed to make the model with menu-dependent attention useful, but we find it difficult to determine a priori what assumptions would be appropriate." In response to this comment, in our model, the menu-dependent consideration probabilities are determined endogenously through the working of peer influence and the structure imposed therein is adequate to both place observable restrictions on the choice data as well as identify the model uniquely. What is it about this endogenizing that produces a very different conclusion with regards to menu-dependent consideration probabilities in our model compared to Manzini and Mariotti (2014)?

To understand this, note that unlike in the case of general menu-dependent consideration probabilities which imposes no restriction on these probabilities, in our setting endogenizing them constrains them in important ways. First, there exists a common positive lower bound on the consideration probabilities of every alternative in any given menu for each

[8]Refer to Theorem 2 in Manzini and Mariotti (2014)

$i \in I$, given by β_i. Further, this bound in menu-independent and imposes meaningful restrictions on behavior. Next, consideration probability of any alternative is increasing in the choice probabilities of others in the cluster. Choice probabilities, in turn, are non-decreasing in consideration probabilities. Together, this produces a complementarity in choice probabilities and generates the sharp restriction that an individual can choose an alternative with probability one if and only if everyone else in her cluster does so. This as we have seen pins down the clusters uniquely in the presence of homophily.

Further, once clusters are pinned down, preferences and the influence coefficients are uniquely identified as well. Manzini and Mariotti (2014) show that for any random choice rule, one can pick any arbitrary set of preferences, and the unconstrained nature of menu-dependent consideration probabilities are such that we can always construct them so as to be able to rationalize the choice rule as a menu-dependent random consideration set rule. However, in our model, such permissibility doesn't exist. Once the clusters are determined, the comparison between an individual's choice probabilities and the average choice probability in their cluster leaves room for only one set of preferences to be compatible with the logic of clustered choices. Particularly, $a \succ_i b$ if and only if $\frac{1-p_i(a, ab)}{1-p_i(b, ab)} \leq \frac{1-\mu_i(a, ab)}{1-\mu_i(b, ab)}$ (see proof of Theorem 4.1). Finally, with preferences determined, consideration probabilities and the influence coefficients that give rise to them are also uniquely pinned down, generating exact identification even in the presence of menu-dependent consideration probabilities.

6.2 Stochastic rationality

For any $i \in I$, given that the smaller is β_i, the greater is peer influence, for influence not to override her preferences, β_i has to be sufficiently high. The following result provides a sufficient condition for choice probabilities to be in line with preferences.

Proposition 6.1. *For any $i \in I$, if $\beta_i \geq \frac{1}{2}$, then for any $A \in \mathcal{X}$ and $a, b \in A$, $a \succ_i b \implies p_i(a, A) \geq p_i(b, A)$, holding strictly whenever $p_i(a, A) > 0$.*[9]

Proof: Please refer to Section A.6.

We can also show that our model violates regularity through a simple example.

Example 6.1. *Consider a cluster with two individuals, $\{1, 2\}$, who agree on $a \succ b$, though $c \succ_1 a$ and $a \succ_2 c$. Then by Lemma A.1, $p_i(a, ab) = 1$ for $i = 1, 2$, implying that $p_i(b, ab) = 0$. However, since $\{a, b, c\}$ is not top-agreeable, $p_i(b, abc) > 0$ for $i = 1, 2$ by the same Lemma. Thus, the model violates regularity.*

[9] Of course, it follows from the result that whenever $\beta_i \geq \frac{1}{2}$, if $p_i(a, A) > p_i(b, A)$ then $a \succ_i b$.

6.3 Status quo bias

One way to interpret the probability assigned to the default alternative in the model is as the likelihood of sticking to a status quo option. As is well known, the status quo or default option plays an important role in many economic problems. In our first example, we look at default probabilities in our model and propose a theory for why a status quo bias may exist in terms of social influence. Because consideration probabilities in our model are endogenous, so are default probabilities. Hence, we can, in terms of our parameters, ask the question as to when are default probabilities high and when are they low? We analyze this here for the case that there is one cluster and a menu with two alternatives. To keep matters simple, we will assume that the influence parameter β for all individuals is the same. We show below that our model then implies that the default probability of all individuals in the cluster is the same.

Let $A = \{a, b\}$, n be the number of people in the cluster and k be the number of people who prefer a over b. Let $p_i(a, ab) = x$, when $a \succ_i b$ and $p_i(b, ab) = y$, when $b \succ_i a$. Then it immediately follows that,

$$p_i(b, ab) = (1 - x)y \text{ if } a \succ_i b$$

$$p_i(a, ab) = (1 - y)x \text{ if } b \succ_i a$$

Hence,

$$x = \beta + (1 - \beta)\left(\frac{kx + (n - k)(1 - y)x}{n}\right), \text{ if } a \succ_i b$$

$$y = \beta + (1 - \beta)\left(\frac{k(1 - x)y + (n - k)y}{n}\right), \text{ if } b \succ_i a$$

Solving these two equations we get,

$$y = \frac{n\beta}{n\beta + k(1 - \beta)x}$$

and

$$x = \left(\sqrt{(n - 2k + 2k\beta)^2 + 4n\beta(1 - \beta)k} - (n - 2k + 2k\beta)\right)\left(\frac{1}{2k(1 - \beta)}\right)$$

The default probability is then given by:

$$p_i(a^*, ab) = (1 - x)(1 - y), \text{ for all } i \in I$$

Proposition 6.2. $\forall i \in I$, *the default probability* $p_i(a^*, ab)$ *increases as* β *decreases. Additionally,* $\forall \beta \in (0, 1)$, *the default probability attains a maxima when* $\frac{k}{n} = \frac{1}{2}$, *i.e., when the population consists of equal proportion of each type.*

Proof: Please refer to Section A.7.

In other words, what the result establishes is that the more susceptible are individuals to influence, the higher is the likelihood of a status quo bias. Further, this likelihood is also higher the more evenly split society is between the two options that are available to choose from.

Another dimension with respect to which status quo bias is observed is when expanding the set of available alternatives. Predominantly, the model predicts a diminishing probability of choosing the default alternative as menu size increases, since the probability of choosing the default in a menu of size M is bounded by $(1-\beta_i)^M$ for each $i \in I$. Since this decreases to 0 as M increases, the probability of choosing the default alternative must vanish asymptotically, irrespective of preferences. However, when adding an alternative causes a cluster to move from a general state of agreement to one of disagreement, the probability of choosing the default alternative may even increase.

A striking example of this is when the addition of an alternative, a, to a menu, $A \in \mathcal{X}$, which is top-agreeable for some cluster N^s, results in a menu, $A \cup a$, which is not top-agreeable for N^s. By Lemma A.1, $\exists b \in A$ s.t. $p_i(b,\, A) = 1 \forall i \in N^s$. Then $p_i(a^*,\, A) = 0 \forall i \in N^s$. However, upon adding a such that $a = \max_{\succ_i} A \cup a$ for some but not all $i \in N^s$, Lemma A.1 implies that $p_i(a',\, A \cup a) > 0$ for all $i \in N^s$, $a' \in A \cup a$. The consequence is that $p_i(a^*,\, A \cup a) > 0$ for all $i \in N^s$.

A Appendix

A.1 Proof of Theorem 3.1

Note, since all interactions are intra-cluster and there are no inter-cluster interactions, it suffices to show that a clustered choice profile exists in every cluster. Secondly, there are no inter-menu interactions, it suffices to show that clustered choices exists for any given menu. Consider a cluster N^t with $|N^t| = J > 1$.

Now consider any menu $A = \{a_1, \ldots, a_M\}$. Let $\Delta_M = \{(q_1, \ldots, q_M) \in \mathbb{R}_+^M : \sum_{m=1}^M q_m \leq 1\}$ denote the unit M-simplex in $\mathbb{R}_+^M$. Define the mapping $\zeta = (\zeta_{i,m})_{i=1,\ldots,J}^{m=1,\ldots,M} : \Delta_M^J \to \Delta_M^J$ as follows: for any $i \in N^t$, $m \in \{1, \ldots, M\}$, and $p = (p_{j,m})_{j=1,\ldots,J}^{m=1,\ldots,M} \in \Delta_M^J$, let

$$\zeta_{i,m}(p) = \left(\beta_i + (1-\beta_i)\frac{\sum_{j=1}^J p_{j,m}}{J}\right) \prod_{m' : a_{m'} \succ_i a_m} (1-\beta_i)\left(1 - \frac{\sum_{j=1}^J p_{j,m'}}{J}\right)$$

We first establish that ζ is a well defined mapping. To do so, establishing the following claim suffices.

Claim: For any $i \in N^t$ and $p \in \Delta_M^J$, $(\zeta_{i,1}(p), \zeta_{i,2}(p), \ldots, \zeta_{i,M}(p)) \in \Delta_M$.

Proof. For each $j \in N^t$ and $m = 1, \ldots, M$, $p_{j,m} \in [0,1]$ and, hence, $\mu_{i,m}(p) := \frac{\sum_{j=1}^{J} p_{j,m}}{J} \in [0,1]$. This in turn implies that $\gamma_{i,m}(p) := \beta_i + (1-\beta_i)\mu_{i,m}(p) \in [0,1]$ as $\beta_i \in (0,1)$; and, hence, $1 - \gamma_{i,m}(p) \in [0,1]$. Accordingly, $\zeta_{i,m}(p) = \gamma_{i,m}(p) \prod_{\{m' : a_{m'} \succ_i a_m\}} (1 - \gamma_{i,m'}(p))$ is a product of terms which all lie in $[0,1]$ and so $\zeta_{i,m}(p) \in [0,1]$. Further, $\zeta_{i,M+1}(p) = \prod_{m=1}^{M}(1 - \gamma_{i,m}(p))$ is also a product of terms which all lie in $[0,1]$ and, hence, $\zeta_{i,M+1}(p) \in [0,1]$ as well.

Next, we establish that $\sum_{m=1}^{M} \zeta_{i,m}(p) \le 1$. It is sufficient to establish that for each p there exists a non-negative number $z(p)$ such that $z_i(p) + \sum_{i=1}^{M} \zeta_{i,\,m}(p) = 1$. To that end, define $z_i : \Delta_M^J \to \mathbb{R}_+$ as

$$z_i(p) = \prod_{m=1}^{M} (1 - \gamma_{i,\,m}(p))$$

Since $\gamma_{i,\,m}(p) \in (0,\,1]$ for each $1 \le m \le M$, $z_i(p) \in [0,\,1]$. Further, assume w.l.o.g. that $a_1 \succ_i a_2 \succ_i \cdots \succ_i a_M$.

By definition, $\zeta_{i,\,M}(p) + z_i(p) = \gamma_{i,\,M} \prod_{m=1}^{M-1}(1-\gamma_{i,\,m}) + \prod_{m=1}^{M}(1-\gamma_{i,\,m}) = \prod_{m=1}^{M-1}(1-\gamma_{i,\,m})$.[10] Further, if $z_i(p) + \sum_{m=\ell}^{M} \zeta_{i,\,m}(p) = \prod_{m=1}^{\ell-1}(1-\gamma_{i,\,m})$ for some $3 \le \ell \le M$, then

$$\begin{aligned} z_i(p) + \sum_{m=\ell-1}^{M} \zeta_{i,\,m}(p) &= \zeta_{i,\,\ell-1}(p) + \prod_{m=1}^{\ell-1}(1-\gamma_{i,\,m}) \\ &= \gamma_{i,\,\ell-1} \prod_{m=1}^{\ell-2}(1-\gamma_{i,\,m}) + (1-\gamma_{i,\,\ell-1}) \prod_{m=1}^{\ell-2}(1-\gamma_{i,\,m}) \\ &= \prod_{m=1}^{\ell-2}(1-\gamma_{i,\,m}) \end{aligned}$$

Accordingly, $z_i(p) + \sum_{m=2}^{M} \zeta_{i,\,m}(p) = 1 - \gamma_{i,\,1}$ and $\zeta_{i,\,1}(p) = \gamma_{i,\,1}$, which means that $z_i(p) + \sum_{m=1}^{M} \zeta_{i,\,m}(p) = 1 - \gamma_{i,\,1} + \gamma_{i,\,1} = 1$. Hence, $\sum_{m=1}^{M} \zeta_{i,\,m}(p) \le 1$ and $(\zeta_{i,1}(p), \ldots, \zeta_{i,M}(p)) \in \Delta_M$ and accordingly the mapping ζ is well defined. $\square$

To show that clustered choices exist in menu A, it suffices to show that ζ has a fixed point. The uniqueness of this fixed point would yield the uniqueness of the clustered choice profile. To that end, we use Banach's fixed point theorem.

Theorem A.1 (Banach's Fixed Point Theorem)**.** *Every contraction mapping from a complete metric space to itself admits a unique fixed point.*

[10] We write $\gamma_{i,m}$ instead of $\gamma_{i,m}(p)$ to economize on notation.

$(\mathbb{R}^{M\times J}, ||\cdot||)$ is a complete metric space, where $||\cdot||$ is a norm on $\mathbb{R}^{M\times J}$. Since Δ_M^J is a compact subset of $\mathbb{R}^{M\times J}$, it is also a complete metric space under the metric induced by the norm. Then, it is enough to show that ζ is a contraction mapping on the open set D_M^J, where $D_M = \text{int}(\Delta_M)$. To do so, write ζ as a composition of maps. First, let μ be the vector of average choice probabilities of each alternative in the cluster. That is, $\mu(m) = \frac{\sum_{j=1}^{J} p_{j,m}}{J}$. Let μ_i denote the vector that permutes the components of μ in order of $\succ_i$. That is, $\mu_{i,m}$ is the average probability of choosing the $\succ_i$ m-th best alternative in A. Then let $\nu : \Delta_M^J \to \Delta_M^J$ be a map such that its i-th component $\nu_i : \Delta_M^J \to \Delta_M$ is defined as $\nu_i(p) = \mu_i$. Now consider the operator $\overline{T} : \mathbb{R}^{M\times J} \to \mathbb{R}^{M\times J}$ that takes a vector $v \in \mathbb{R}^{M\times J}$ and is defined as $\overline{T}_{m+jM}(v) = \overline{T}_m(v) = J^{-1}\sum_{j=0}^{J-1} v_{m+jM}$. When acting on $p \in \Delta_M^J$, $\overline{T}$ gives J vertically stacked copies of μ. Since ν is a permuted version of $\overline{T}$, $|\nu| = |\overline{T}|$, where $|\cdot|$ is the operator norm, defined for a linear operator $T : \mathbb{R}^{M\times J} \to \mathbb{R}^{M\times J}$ as $|T| = \sup_{v\in\mathbb{R}^{M\times J}} \frac{||Tv||}{||v||}$. Since $\overline{T}$'s matrix under the standard basis is $J^{-1}\cdot \mathbf{1}_{J\times J} \otimes I_{M\times M}$,[11] which is a symmetric idempotent matrix, $|\overline{T}| = 1$. Thus, $|\nu| = 1$ and ν is a non-expansive mapping.

Now, consider the mapping $\eta : \Delta_M^J \to \Delta_M^J$, such that each of its components $\eta_i : \Delta_M \to \Delta_M$ take a vector μ_i and yield a new vector of choice probabilities $\eta_i(\mu_i)$, such that applying a permutation mapping $P_i : \Delta_M \to \Delta_M$ that reorders choice probabilities in the original order of $A = \{a_1, ..., a_M\}$, to satisfy $P_i(\eta_i(\mu_i)) = \zeta_i(p)$.

To see that η is indeed a contraction mapping on Δ_M^J, we start by looking at the Jacobian matrix of η_i, $\partial\eta_i$, and start by showing that

$$\sup_{D_M} \rho(\partial\eta_i(\mu_i)) < 1$$

where $\rho(T)$ denotes the spectral radius (largest eigenvalue in absolute value) of an operator T. To do so, note that $\partial\eta_i(\mu_i)$ is a lower triangular matrix, and its eigenvalues are its diagonal entries, which are

$$\frac{\partial \eta_{i,m}}{\partial \mu_{i,m}}\Big|_{\mu_i} = (1-\beta_i) \prod_{m':a_{m'}\succ_i a_m} (1-\beta_i)(1-\mu_{i,m'})$$

Since each of the terms $(1-\beta_i)(1-\mu_{i,m'}) < 1$, $\rho(\partial\eta_i(\mu_i)) = 1-\beta_i$. Note then that $\partial\eta(\mu)$ for $\mu \in D_M^J$ is simply the block diagonal matrix with each block being $\partial\eta_i(\mu_i)$. This implies that $\partial\eta(\mu)$ is also lower triangular, and $\sup_{D_M^J} \rho(\partial\eta(\mu)) = \max_j \sup_{D_M} \rho(\partial\eta_j(\mu_j)) = 1 - \min_j \beta_j < 1$. Then, η is a locally contractive mapping on all of Δ_M^J (Hefti, 2015). Note, since Δ_M^J is compact, η is uniformly locally contractive (Jungck, 1982). Then, since Δ_M^J is convex, the shortest path between any two points in Δ_M^J is given by a straight line between them. By Lemma 2.2 of Ciesielski and Jasinski (2016), η is a contraction mapping

[11] $\mathbf{1}_{J\times J}$ is a square matrix of ones, $I_{M\times M}$ is the identity matrix of size M, and $\otimes$ is the Kronecker product

on Δ_M^J.[12] Then, by the non-expansiveness of ν, and $P = (P_j)_{j=1}^J$ being a permutation map, $\zeta : P \circ \eta \circ \nu$ is a contraction map on Δ_M^J. Then, by the Banach fixed point theorem, ζ has a unique fixed point p^*.

A.2 Key Lemmas

We now prove a few lemmas that hold for a clustered choice profile. These lemmas will be used to prove the identification result (Theorem 4.1) and the necessity of the axioms for the representation (Theorem 5.1). Let $(p_i)_{i\in I}$ have a clustered choice representation $\left\langle \{\succ_i \colon i \in I\}, N = \{N^1, \dots, N^S\}, \{\beta_i : i \in I\} \right\rangle$. In the way of notation, for any $i \in I$, $A \in \mathcal{X}$ and $a \in A$, let

$$\hat{\mu}_{N(i)}(a, A) = \frac{1}{|N(i)|} \sum_{j \in N(i)} p_j(a, A)$$

denote the average probability of choosing a in A in i's cluster $N(i)$. Whenever there is no ambiguity regarding the underlying clustering we are referring to when evaluating this average choice probability, we will simply write $\hat{\mu}_i$ instead of $\hat{\mu}_{N(i)}$. Further, we say that a non-singleton menu $A \in \mathcal{X}$ is a *full support menu (FSM)* for $i \in I$ if $p_i(a, A) > 0$ for all $a \in A$. We denote the set of all FSMs for i by $\mathcal{X}_i$. Also, recall that we refer to a menu A as top-agreeable for a cluster $N^s \in N$ if the $\succ_i$-best alternative in A for all $i \in N^s$ is the same. Then, the following conclusion follows.

Lemma A.1. *For every $i \in I$, A is not top agreeable for i's cluster $N(i)$ iff $A \in \mathcal{X}_i$.*

Proof. Suppose A is not top agreeable for $N(i)$. Then, if $a = \max_{\succ_i} A$, $\exists j \in N(i)$ s.t. $b = \max_{\succ_j} A$, $a \neq b$. Since $\beta_k > 0$, $\forall\, k \in I$, $p_i(a, A) = \gamma_i(a, A) = \beta_i + (1-\beta_i)\hat{\mu}_i(a, A) > 0$, and $p_j(b, A) = \gamma_j(b, A) = \beta_j + (1-\beta_j)\hat{\mu}_i(b, A) > 0$. This implies $p_i(a', A) < 1,\ \forall a' \in A \setminus a;\ p_j(a', A) < 1, \forall a' \in A \setminus b$. That is, $\forall a' \in A,\ p_\ell(a', A) < 1$, for some $\ell \in N(i)$. Accordingly, $\hat{\mu}_i(a', A) < 1,\ \forall a' \in A$.

Now consider any $c \in A$. Since $\hat{\mu}_i(c, A) < 1$ and $\beta_i \in (0, 1)$, $\gamma_i(c, A) = \beta_i + (1 - \beta_i)\hat{\mu}_i(c, A) \in (0, 1)$. $\gamma_i(c, A) \in (0, 1)$ implies $1 - \gamma_i(c, A) \in (0, 1)$. As a consequence, $p_i(c, A) = \gamma_i(c, A) \prod_{c' \succ_i c}(1 - \gamma_i(c', A)) \in (0, 1)$. Since this is true for all $c \in A$, $A \in \mathcal{X}_i$.

Next, suppose A is top agreeable for $N(i)$, with a being the $\succ_j$-best alternative in A for all $j \in N(i)$. Suppose $p_i(a, A) < 1$. Then $\hat{\mu}_i(a, A) < 1$. For all $j \in N(i)$, $\beta_j > 0$ implies

$$p_j(a, A) = \gamma_j(a, A) = \beta_j + (1 - \beta_j)\hat{\mu}_i(a, A) > \hat{\mu}_i(a, A)$$

[12]Formally, Δ_M^J is connected, compact, and convex. As a result, it is ϵ-chainable in the language of Ciesielski and Jasinski (2016), where the $D_\epsilon(x, y)$, defined as the length of the shortest ϵ-chain from x to y is equal to $d(x, y)$ (d being the metric induced by norm $||\cdot||$ on Δ_M^J) due to convexity. As such, Δ_M^J is a contraction with respect to d.

That is, $\forall j \in N(i), p_j(a, A) > \hat{\mu}_i(a, A)$, which implies $\hat{\mu}_i(a, A) > \hat{\mu}_i(a, A)$! Hence, under a clustered choice profile, $p_i(a, A) = 1$ and $A \notin \mathcal{X}_i$. $\square$

The proof of the Lemma establishes the following corollaries.

Corollary A.1. *The following statements are equivalent:*

1. $A \notin \mathcal{X}_i$, $i \in I$
2. $p_i(a, A) = 1$, *for some* $a \in A$
3. *A is top agreeable for $N(i)$ with a the $\succ_j$-best alternative in* $A, \forall j \in N(i)$
4. $p_j(a, A) = 1, \forall j \in N(i)$

Corollary A.2. *Let* $\left\langle \{\succ_i : i \in I\}, N = \left\{N^1, ..., N^S\right\}, \{\beta_i : i \in I\}\right\rangle$ *be a clustered choice representation of the profile of non-trivial random choice rules* $(p_i)_{i \in I}$. *For each cluster* $N^s \in N$, *there exists* $A \in \mathcal{X}$ *such that A is not top-agreeable for* N^s.

Proof. Suppose there exists a cluster $N^s \in N$ such that every menu in $\mathcal{X}$ is top-agreeable for N^s. It then follows from the conclusion in Lemma A.1 and Corollary A.1 that for each $A \in \mathcal{X}$, there exists $a \in A$ such that $p_i(a, A) = 1$ for all $i \in N^s$. However, since this is true for all A, it means that the random choice rule p_i is not non-trivial for any $i \in N^s$! Accordingly, there exists $A \in \mathcal{X}$ such that A is not top-agreeable for N^s. $\square$

Lemma A.2. *Let* $A \in \mathcal{X}$ *and* $i \in I$. *Then:*

(i) $A \in \mathcal{X}_i \implies \left[a = \max_{\succ_i} A \iff \frac{1-p_i(a,A)}{1-p_i(b,A)} < \frac{1-\hat{\mu}_i(a,A)}{1-\hat{\mu}_i(b,A)}, \forall b \in A \setminus a\right]$

(ii) $A \notin \mathcal{X}_i \implies [a = \max_{\succ_i} A \iff p_i(a, A) = 1]$

Proof. Let $A = \{a_1, \ldots, a_M\} \in \mathcal{X}_i$ be s.t. $a_1 \succ_i a_2 \succ_i \cdots \succ_i a_M$. Since $A \in \mathcal{X}_i$, for the derivation below, note that $\sum_{m=1}^{l} p_i(a_m, A) \in (0, 1)$, or equivalently, $1 - \sum_{m=1}^{l} p_i(a_m, A) \in (0, 1)$ for all $l = 1, \ldots, M - 1$. Now,

$$\begin{aligned} p_i(a_1, A) &= \gamma_i(a_1, A) = \beta_i + (1 - \beta_i)\hat{\mu}_i(a_1, A) \\ \Rightarrow 1 - \beta_i &= \frac{1 - p_i(a_1, A)}{1 - \hat{\mu}_i(a_1, A)} \end{aligned}$$

Note that, under a clustered choice profile, an alternative $a_l \in A$ not being chosen corresponds to the event that either this alternative is not considered or some alternative preferred to it is considered. Accordingly, it is straightforward to derive that the probability that the m $\succ_i$-top alternatives are not chosen, $1 - \sum_{k \leq m} p_i(a_k, A)$, is equal to the probability that none of them is considered, which by the independence of the consideration of alternatives is $\prod_{k\leq m}(1-\gamma_i(a_k,\, A))$. Then,

$$\begin{aligned} p_i(a_m,\, A) &= \gamma_i(a_m,\, A) \prod_{k\leq m-1}(1-\gamma_i(a_k,\, A)) \\ \Longrightarrow\; \gamma_i(a_m,\, A) &= \frac{p_i(a_m,\, A)}{1-\sum_{k\leq m-1} p_i(a_k,\, A)} \\ \Longrightarrow\; 1-[\beta_i+(1-\beta_i)\hat{\mu}_i(a_m,\, A)] &= 1-\frac{p_i(a_m,\, A)}{1-\sum_{k\leq m-1} p_i(a_k,\, A)} \end{aligned}$$

Accordingly,

$$1-\beta_i = \frac{1-\frac{p_i(a_m,\, A)}{1-\sum_{k\leq m-1} p_i(a_k,\, A)}}{1-\hat{\mu}_i(a_m,\, A)} \tag{3}$$

Since $1-\sum_{k\leq m-1} p_i(a_k,\, A) \in (0,\, 1)$ for all $2 \leq m \leq M$,

$$\begin{aligned} 1-\frac{p_i(a_m,\, A)}{1-\sum_{k\leq m-1} p_i(a_k,\, A)} &< 1-p_i(a_m,\, A) \\ \Longrightarrow\; \frac{1-p_i(a_1,\, A)}{1-\hat{\mu}_i(a_1,\, A)} = 1-\beta_i &< \frac{1-p_i(a_m,\, A)}{1-\hat{\mu}_i(a_m,\, A)} \end{aligned}$$

Hence, $\frac{1-p_i(a,A)}{1-\hat{\mu}_i(a,A)} < \frac{1-p_i(b,A)}{1-\hat{\mu}_i(b,A)}$ or $\frac{1-p_i(a,A)}{1-p_i(b,A)} < \frac{1-\hat{\mu}_i(a,A)}{1-\hat{\mu}_i(b,A)}$, $\forall b \in A \setminus a$, if $a = \max_{\succ_i} A$. To establish the only if direction, suppose $\frac{1-p_i(a,A)}{1-p_i(b,A)} < \frac{1-\hat{\mu}_i(a,A)}{1-\hat{\mu}_i(b,A)}$, $\forall b \in A \setminus a$, but $\hat{a} \neq a$ is $\succ_i$-best in A. Then based on the argument above, it follows that $\frac{1-p_i(\hat{a},A)}{1-p_i(a,A)} < \frac{1-\hat{\mu}_i(\hat{a},A)}{1-\hat{\mu}_i(a,A)}$, or $\frac{1-p_i(a,A)}{1-p_i(\hat{a},A)} > \frac{1-\hat{\mu}_i(a,A)}{1-\hat{\mu}_i(\hat{a},A)}$, a contradiction!

Now consider the case $A \notin \mathcal{X}_i$. Lemma A.1 implies that A is top agreeable for $N(i)$. Then as shown in the proof of that Lemma and stated in Corollary A.1, $a = \max_{\succ_i} A$ iff $p_i(a, A) = 1$. □

Lemma A.2 implies the following corollary:

Corollary A.3. *For all $i \in I$ and $A \in \mathcal{X}$,*

$$a = \max_{\succ_i} A \iff \frac{1-p_i(a,A)}{1-p_i(b,A)} \leq \frac{1-\hat{\mu}_i(a,A)}{1-\hat{\mu}_i(b,A)}, \forall b \in A \setminus a$$

Proof. If $A \in \mathcal{X}_i$, then the conclusion follows immediately from the first statement of Lemma A.2.

Now suppose $A \notin \mathcal{X}_i$. By the second statement of Lemma A.2, $a = \max_{\succ_i} A$ iff $p_i(a,\, A) = 1$ iff $p_i(b,\, A) = 0$, $\forall b \in A \setminus a$. Then, for any $b \in A \setminus a$, $\frac{1-p_i(a,\, A)}{1-p_i(b,\, A)} = \frac{0}{1} = 0$. Further,

$p_i(b,\,A)=0 \implies \hat{\mu}_i(b,\,A)<1$. Accordingly, $\frac{1-\hat{\mu}_i(a,A)}{1-\hat{\mu}_i(b,A)}$ is defined and non-negative. Thus, $\frac{1-p_i(a,A)}{1-p_i(b,A)} \leq \frac{1-\hat{\mu}_i(a,A)}{1-\hat{\mu}_i(b,A)}, \forall b \in A \setminus a$. To establish the other direction, suppose $\frac{1-p_i(a,A)}{1-p_i(b,A)} \leq \frac{1-\hat{\mu}_i(a,A)}{1-\hat{\mu}_i(b,A)}$, $\forall\, b \in A \setminus a$, but $a' = \max_{\succ_i} A \neq a$. Then, by Lemma A.2, $1-p_i(a',\,A)=0$ and $\frac{1-p_i(a,A)}{1-p_i(a',A)}$ is undefined. Thus, the inequality cannot hold for all $b \in A \setminus a$, leading us to a contradiction. $\square$

The following conclusion follows immediately from the last result.

Corollary A.4. *For all $i \in I$ and $a, b \in X$,*

$$a \succ_i b \iff \frac{1-p_i(a,ab)}{1-p_i(b,ab)} \leq \frac{1-\hat{\mu}_i(a,ab)}{1-\hat{\mu}_i(b,ab)}$$

A.3 Proof of Theorem 4.1

Let $\left\langle \{\succ_i : i \in I\}, N = \{N^1,\ldots,N^S\}, \{\beta_i : i \in I\} \right\rangle$ and $\left\langle \{\hat{\succ}_i : i \in I\}, \hat{N} = \{\hat{N}^1,\ldots,\hat{N}^T\}, \{\hat{\beta}_i : i \in I\} \right\rangle$ be two representations of the clustered choice profile $(p_i)_{i\in I}$.

For each $i \in I$, define $R(i)$ as

$$R(i) = \{j \in I : p_i(a,\,A)=1 \iff p_j(a,\,A)=1,\, A \in \mathcal{X},\, a \in A\}$$

<u>**Uniqueness of Clusters**</u>: By Corollary A.1, $p_i(a,\,A)=1 \iff p_j(a,\,A)=1 \forall j \in N(i)$. Thus, $j \in N(i) \implies j \in R(i)$. Now suppose $j \notin N(i)$. Since the clustering is homophilous, there exists a menu A such that A is top agreeable for one of $N(i)$ and $N(j)$, and not for the other with the same top. Suppose it is top agreeable for $N(i)$ with top $a \in A$. Then there exists $j' \in N(j)$ such that $a \neq \max_{j'} A$. By Corollary A.1, $p_i(a,\,A)=1$ and $p_j(a,\,A)\neq 1$. Thus, $j \notin R(i)$. Likewise, if A is top agreeable for $N(j)$ with top a, then $\exists i' \in N(i)$ such that $a \neq \max_{i'} A$. By Corollary A.1, $p_j(a,\,A)=1$ but $p_i(a,\,A)\neq 1$, which also implies $j \notin R(i)$. Then $j \notin N(i) \implies j \notin R(i)$, which means $R(i) \subseteq N(i)$. Thus, $N(i) = R(i)$ for all $i \in I$.

Since the same argument applies to $\hat{N}$, $N(i) = R(i) = \hat{N}(i)$ for all $i \in I$, which implies that $\hat{N} = N$.

<u>**Uniqueness of Preferences**</u>: Since $N(i) = \hat{N}(i)$, $\hat{\mu}_{N(i)} = \hat{\mu}_{\hat{N}(i)}$. Then, by Corollary A.4,

$$a \succ_i b \iff \frac{1-p_i(a,\,ab)}{1-p_i(b,\,ab)} \leq \frac{1-\hat{\mu}_{N(i)}(a,\,ab)}{1-\hat{\mu}_{N(i)}(b,\,ab)} = \frac{1-\hat{\mu}_{\hat{N}(i)}(a,\,ab)}{1-\hat{\mu}_{\hat{N}(i)}(b,\,ab)} \iff a \hat{\succ}_i b$$

Thus, $\succ_i = \hat{\succ}_i$ for all $i \in I$.

Uniqueness of β_i: We know from Corollary A.2 that there exists $A \in \mathcal{X}$ such that A is not top-agreeable for the cluster $N(i)$ under the first representation. Since $N(i) = \hat{N}(i)$ and $\succ_i = \hat{\succ}_i$, it follows that the menu A is also not top-agreeable for the cluster $\hat{N}(i)$ under the second representation. Let $a = \max_{\succ_i} A = \max_{\hat{\succ}_i} A$. We also know from Corollary A.1 that $A \in \mathcal{X}_j$, for all $j \in N(i) = \hat{N}(i)$. Accordingly, $p_i(a, A) \in (0, 1)$ and $\hat{\mu}_{N(i)}(a,\, A) = \hat{\mu}_{\hat{N}(i)}(a,\, A) \in (0, 1)$. Putting everything together, we have

$$\begin{aligned} p_i(a,\, A) &= \beta_i + (1 - \beta_i)\hat{\mu}_{N(i)}(a,\, A) \\ &= \hat{\beta}_i + (1 - \hat{\beta}_i)\hat{\mu}_{\hat{N}(i)}(a,\, A) \\ \implies \beta_i &= 1 - \frac{1 - p_i(a,\, A)}{1 - \hat{\mu}_{N(i)}(a,\, A)} = \frac{1 - p_i(a,\, A)}{1 - \hat{\mu}_{\hat{N}(i)}(a,\, A)} = \hat{\beta}_i \end{aligned}$$

Thus, $\beta_i = \hat{\beta}_i$ for all $i \in I$.

A.4 Proof of Proposition 4.1

Let $\mathcal{P}$ be the set of all clustered choice profiles $(\tilde{p}_i)_{i \in I}$ such that $\tilde{p}_i = p_i$ on $\bar{\mathcal{X}}$. Let $\left\langle \{\succ_i : i \in I\}, N = \{N^1, \ldots, N^S\}, \{\beta_i : i \in I\} \right\rangle$ and $\left\langle \{\hat{\succ}_i : i \in I\}, \hat{N} = \{\hat{N}^1, \ldots, \hat{N}^T\}, \{\hat{\beta}_i : i \in I\} \right\rangle$ be the parameters associated with any two clustered choice profiles in $\mathcal{P}$.[13]

Identifying Clusters: By Corollary A.1, $A \notin \mathcal{X}_i$ iff A is top agreeable for $N(i)$ with some $a \in A$ the $\succ_j$-best alternative in A for all $j \in N(i)$. This is true iff $\{a, b\} \subseteq A$ is top-agreeable for $N(i)$ with $a \succ_j b$, for all $b \in A \setminus a$, $j \in N(i)$. By Corollary A.1, this holds iff $p_j(a,\, ab) = 1$ for all $b \in A \setminus a$, $j \in N(i)$.

Define $R_{\tilde{p}}(i)$ for $(\tilde{p}_i)_{i \in I} \in \mathcal{P}$ as

$$R_{\tilde{p}}(i) \equiv \{j \in I : \tilde{p}_i(a,\, A) = 1 \iff \tilde{p}_j(a,\, A) = 1, A \in \mathcal{X},\, a \in A\}$$

Further, define $R^*(i)$ as

$$R^*(i) \equiv \left\{j \in I : p_i(a,\, A) = 1 \iff p_j(a,\, A) = 1,\, A \in \bar{\mathcal{X}},\, a \in A\right\}$$

Since $\tilde{p}_i = p_i$ on $\bar{\mathcal{X}}$, it is obvious that $R_{\tilde{p}}(i) \subseteq R^*(i)$. If we can show that $R_{\tilde{p}}(i) = R^*(i)$ for all $(\tilde{p}_i)_{i \in I} \in \mathcal{P}$, then $N = \hat{N}$ by the proof of Theorem 4.1.

To that end, suppose $j \notin R_{\tilde{p}}(i)$. Then there exists $A \in \mathcal{X}$ such that $A \in \mathcal{X}_i$ but $A \notin \mathcal{X}_j$ or vice-versa. Suppose, w.l.o.g. that $A \in \mathcal{X}_i$. Then there exists $i' \in R(i)$ such that $b \succ_{i'} a$ for some $b \in A \setminus a$. Therefore, $\tilde{p}_{i'}(a,\, ab) = p_{i'}(a,\, ab) \neq 1$, implying that $p_i(a,\, ab) \neq 1$.

[13] There can only be one set of parameters for each clustered choice profile in $\mathcal{P}$ by Theorem 4.1.

However, $A \notin \mathcal{X}_j$ iff there exists $a \in A$ such that $\tilde{p}_j(a, A) = 1$ iff $\tilde{p}_j(a, ab) = 1$ for all $b \in A \setminus a$ iff $p_j(a, ab) = 1$ for all $b \in A \setminus a$. However, since $p_i(a, ab) \neq 1$ for some $b \in A \setminus a$, $j \notin R^*(i)$. Thus, $R^*(i) \subseteq R_{\tilde{p}}(i)$.

Since this holds true for all $(\tilde{p}_i)_{i \in I}$, $N(i) = \hat{N}(i) = R^*(i)$ for all $i \in I$.

<u>**Identifying Preferences**</u>: Since $N(i) = \hat{N}(i)$, by the proof of Theorem 4.1, $\succ_i = \hat{\succ}_i$ for all $i \in I$.

<u>**Identifying** β_i</u>: First, we need to establish that there exists a menu $\{a, b\} \in \bar{\mathcal{X}}$ for each $i \in I$ such that $\{a, b\} \in \mathcal{X}_i$. Suppose toward a contradiction that this does not hold. Then the same holds true for $R^*(i) = N(i) = \hat{N}(i)$. By Corollary A.1, $p_i(a, ab) = 1$ implies $a \succ_j b$ for all $j \in R^*(i)$. However, since this is true for all $\{a, b\} \in \bar{\mathcal{X}}$, $\succ_i = \succ_j$ for all $j \in R^*(i)$. However, this would imply that p_i is a trivial random choice rule. Thus, for each $i \in I$, there exists $\{a, b\} \subseteq \bar{\mathcal{X}}$ such that $\{a, b\} \in \mathcal{X}_i$ iff $\{a, b\} \in \mathcal{X}_j$ for all $j \in R^*(i)$. Suppose $\{a, b\} \in \mathcal{X}_i$ and $a \succ_i b$. By the proof of Theorem 4.1,

$$\beta_i = 1 - \frac{1 - p_i(a, ab)}{1 - \hat{\mu}_{R^*(i)}(a, ab)} = \hat{\beta}_i$$

Thus, $\beta_i = \hat{\beta}_i$ for all $i \in I$.

A.5 Proof of Theorem 5.1

Necessity: Suppose $(p_i)_{i \in I}$ is a clustered choice profile with homophilous clustering $N = \{N^1, \ldots, N^S\}$, collection of strict preference rankings $\{\succ_i : i \in I\}$ and influence coefficients $\{\beta_i : i \in I\}$. Then $(p_i)_{i \in I}$ satisfies:

Connectedness: From Corollary A.1, for any menu A, $a \in A$, $p_i(a, A) = 1 \iff A$ is top-agreeable for $N(i)$ with top $a \iff p_j(a, A) = 1, \forall j \in N(i)$. Then any $j \in N(i)$ is connected to i. Further, since $|N(i)| > 1$, there exists $j \in N(i)$, $j \neq i$, such that j and i are connected.

Stochastic IIA: We have established in the proof of Theorem 4.1 that for any clustered choice profile with homophilous clustering $N = \{N^1, \ldots, N^S\}$, $N(i) = R(i)$, for any $i \in I$. Accordingly, $\mu_i(a, A) = \hat{\mu}_i(a, A)$, for any $A \in \mathcal{X}$. Consider any such menu A. Given that $\succ_i$ is a ranking, there exists a unique $a \in A$ such that $a = \max_{\succ_i} A$. By Corollary A.3, it follows that $\frac{1-p_i(a,A)}{1-p_i(b,A)} \leq \frac{1-\hat{\mu}_i(a,A)}{1-\hat{\mu}_i(b,A)}, \forall b \in A \setminus a$. Further, since $\hat{\mu}_i = \mu_i$, we can conclude that $\exists!\ a \in A$ such that $\frac{1-p_i(a,A)}{1-p_i(b,A)} \leq \frac{1-\mu_i(a,A)}{1-\mu_i(b,A)}, \forall b \in A \setminus a$.

Next, consider any $B \subseteq A$ with $a \in B$. Clearly $a = \max_{\succ_i} B$, and it follows from Corollary A.3 and $\hat{\mu}_i = \mu_i$ that $\frac{1-p_i(a,B)}{1-p_i(b,B)} \leq \frac{1-\mu_i(a,B)}{1-\mu_i(b,B)}, \forall b \in B \setminus a$.

Menu Independence of Influence: Let $A = \{a_1, \dots, a_M\} \in \mathcal{X}$ be a menu in which i chooses stochastically. Suppose $a_1 \succ_i a_2 \succ_i \cdots \succ_i a_M$. Then $\bar{A}_i(a_1) = \emptyset$ and $\bar{A}_i(a_m) = \{a_1, \dots, a_{m-1}\}$, for $m = 2, \dots, M$. This is because by Corollary A.1 and the fact that $\mu_i(.) = \hat{\mu}_i(.)$, we know that for any $a, b \in A$, $\frac{1-p_i(a,ab)}{1-p_i(b,ab)} \leq \frac{1-\mu_i(a,ab)}{1-\mu_i(b,ab)}$ iff $a \succ_i b$. Accordingly, $p_i^*(a_1, A) = p_i(a_1, A)$, and

$$p_i^*(a_m, A) = \frac{p_i(a_m, A)}{1 - p_i(a_1, A) - p_i(a_2, A) - \cdots - p_i(a_{m-1}, A)}, \forall m = 2, \dots, M$$

Further, drawing on equation (3) that we established in the proof of Lemma A.2 and the fact that $\mu_i(.) = \hat{\mu}_i(.)$, it follows that for any such A,

$$\frac{1 - p_i(a_1, A)}{1 - \mu_i(a_1, A)} = \frac{1 - \frac{p_i(a_2,A)}{1-p_i(a_1,A)}}{1 - \mu_i(a_2, A)} = \cdots = \frac{1 - \frac{p_i(a_M,A)}{1-p_i(a_1,A)-p_i(a_2,A)-\cdots-p_i(a_{M-1},A)}}{1 - \mu_i(a_M, A)} = 1 - \beta_i$$

i.e.,

$$\frac{1 - p_i(a_1, A)}{1 - \mu_i(a_1, A)} = \frac{1 - p_i^*(a_1, A)}{1 - \mu_i(a_1, A)} = \frac{1 - p_i^*(a_2, A)}{1 - \mu_i(a_2, A)} = \cdots = \frac{1 - p_i^*(a_M, A)}{1 - \mu_i(a_M, A)} = 1 - \beta_i < 1$$

Accordingly, for any $A, B \in \mathcal{X}$ in which i chooses stochastically, and any $a \in A$, $b \in B$, we have

$$\frac{1 - p_i^*(a, A)}{1 - \mu_i(a, A)} = \frac{1 - p_i^*(b, B)}{1 - \mu_i(b, B)}$$

Sufficiency: Let $(p_i)_{i \in I}$ be a profile of non-trivial random choice rules that satisfies connectedness, stochastic IIA, and menu independence of influence. To show that $(p_i)_{i \in I}$ is a clustered choice profile we need to define three things: a homophilous clustering $N = \{N^1, \dots, N^S\}$ of I; a collection of strict preference rankings $\{\succ_i : i \in I\}$; and influence coefficients $\{\beta_i : i \in I\}$.

Defining the clustering: For any $i \in I$, define

$$N(i) = R(i) = \{j \in I : i \text{ and } j \text{ are connected}\}$$

We want to show that this produces a valid partition. First, note that by the symmetry in the definition, $i \in R(j) \iff j \in R(i)$. If $j \in R(i)$, then $p_{j'}(a,\, A) = 1 \iff p_j(a,\, A) = 1 \iff p_i(a,\, A) = 1 \iff p_{i'}(a,\, A) = 1$, for $j' \in R(j)$ and $i' \in R(i)$. That is, $R(i) = R(j)$. By connectedness, $\exists j \neq i$ s.t. $j \in R(i)$, implying that $|N(i)| > 1$.

Finally, suppose $R(i) \neq R(j)$. By the above argument, $j \notin R(i)$. Furthermore, if $j' \in R(i) \cap R(j)$, by the same argument, $R(i) = R(j') = R(j)$, which is a contradiction. Trivially, $\cup_{i \in I} R(i) = I$, which implies that $\{N(i)\}_{i \in I}$ is a valid clustering.

We next define preferences, and show at the end of the proof that this definition is consistent with N being a homophilous clustering.

Defininining preferences: Consider $i \in I$ and define $\succ_i \subseteq X \times X$ by: for any $a, b \in X, a \neq b, a \succ_i b$ if $\frac{1-p_i(a,ab)}{1-p_i(b,ab)} \leq \frac{1-\mu_i(a,ab)}{1-\mu_i(b,ab)}$. First, we establish that $\succ_i$ is total. This follows from stochastic IIA, since for menu $\{a, b\}$, either $\frac{1-p_i(a,ab)}{1-p_i(b,ab)} \leq \frac{1-\mu_i(a,ab)}{1-\mu_i(b,ab)}$ or $\frac{1-p_i(b,ab)}{1-p_i(a,ab)} \leq \frac{1-\mu_i(b,ab)}{1-\mu_i(a,ab)}$. Hence, we have $a \succ_i b$ or $b \succ_i a$. Since, by stochastic IIA, this inequality holds for a unique alternative, it establishes that $\succ_i$ is asymmetric. Finally, to show that $\succ_i$ is transitive, let $a \succ_i b, b \succ_i c$ and consider $A = \{a, b, c\}$. By stochastic IIA, $\exists! \, d \in A$ such that $\frac{1-p_i(d,A)}{1-p_i(e,A)} \leq \frac{1-\mu_i(d,A)}{1-\mu_i(e,A)}$ and $\frac{1-p_i(d,de)}{1-p_i(e,de)} \leq \frac{1-\mu_i(d,de)}{1-\mu_i(e,de)}$, for all $e \in A \setminus d$. By the way $\succ_i$ is defined, given that $a \succ_i b$ and $b \succ_i c$, $d \neq b, c$. Hence, $d = a$, and consequently $a \succ_i c$.

Defining $\beta_i's$: For any $i \in I$, define β_i as follows by taking any menu $A \in \mathcal{X}$ in which i chooses stochastically, and some $a \in A$:

$$\beta_i = \frac{p_i^*(a, A) - \mu_i(a, A)}{1 - \mu_i(a, A)}$$

Since $1 - \beta_i = \frac{1-p_i^*(a,A)}{1-\mu_i(a,A)}$, menu independence of influence guarantees that the definition of β_i is independent of the choice of a and A, and that $1 - \beta_i < 1$, or $\beta_i > 0$. By how preferences are defined, $b \succ_i a$ iff $b \in \bar{A}_i(a)$, $a, b \in A$. Then, for $a' = \max_{\succ_i} A$, since $\bar{A}_i(a') = \emptyset$, $p_i^*(a', A) = \frac{p_i(a', A)}{1-\sum_{a \in \bar{A}_i(a')} p_i(a, A)} = p_i(a', A)$. Since i chooses stochastically in A, $p_i(a', A) < 1$, which implies $\beta_i = \frac{p_i(a', A) - \mu_i(a', A)}{1-\mu_i(a', A)} < 1$.

Establishing the representation: Consider any $i \in I$. If i chooses non-stochastically in A, then there exists $a \in A$ such that $p_i(a, A) = 1 \iff p_j(a, A) = 1$ for all $j \in R(i) = N(i)$. Then:

$$\begin{aligned}
\gamma_i(a, A) &=: \beta_i + (1 - \beta_i) \frac{1}{|N(i)|} \sum_{j \in N(i)} p_j(a, A) \\
&= \beta_i + 1 - \beta_i \\
&= 1 \\
&= p_i(a, A)
\end{aligned}$$

Note that $\frac{1-p_i(a,A)}{1-p_i(b,A)} = 0 = \frac{1-\mu_i(a,A)}{1-\mu_i(b,A)}$, for all $b \in A \setminus a$. By stochastic IIA, this implies that $\frac{1-p_i(a,ab)}{1-p_i(b,ab)} \leq \frac{1-\mu_i(a,ab)}{1-\mu_i(b,ab)}$, for all $b \in A \setminus a$, which implies by the definition of $\succ_i$ that $a = \max_{\succ_i} A$. Then, for all $b \in A \setminus a$

$$\begin{aligned}
\gamma_i(b, A) \prod_{c \succ_i b;\, c \in A} (1 - \gamma_i(c, A)) &= \gamma_i(b, A)(1 - \gamma_i(a, A)) \prod_{c \succ_i b; c \in A \setminus a} (1 - \gamma_i(c, A)) \\
&= 0 \\
&= p_i(b, A)
\end{aligned}$$

Hence, the representation holds for such A.

Now, consider $A = \{a_1, ..., a_M\}$ in which i chooses stochastically, and assume wlog that $a_1 \succ_i a_2 \succ_i ... \succ_i a_M$. By the definition of $\succ_i$, $\bar{A}_i(a_1) = \emptyset$ and $\bar{A}_i(a_m) = \{a_1, ..., a_{m-1}\}$ for $m \in \{2, ..., M\}$. By menu independence of influence,

$$\beta_i = \frac{p_i^*(a_1, A) - \mu_i(a_1, A)}{1 - \mu_i(a_1, A)} = \frac{p_i(a_1, A) - \mu_i(a_1, A)}{1 - \mu_i(a_1, A)},$$

which implies

$$p_i(a_1, A) = \beta_i + (1 - \beta_i)\mu_i(a_1, A) =: \gamma_i(a_1, A)$$

Again applying menu independence of influence, we have that for any a_m,

$$\begin{aligned} &\frac{p_i^*(a_m, A) - \mu_i(a_m, A)}{1 - \mu_i(a_m, A)} = \beta_i \\ &\qquad \implies p_i^*(a_m, A) = \beta_i + (1 - \beta_i)\mu_i(a_m, A) =: \gamma_i(a_m, A) \end{aligned}$$

Accordingly,

$$\frac{p_i(a_m, A)}{1 - \sum_{k \le m-1} p_i(a_k, A)} = \gamma_i(a_m, A) \tag{4}$$

Next, we establish by induction that $1 - \sum_{k \le \ell} p_i(a_k, A) = \prod_{k \le \ell}(1 - \gamma_i(a_k, A))$. To do so, first, we know from above that $1 - p_i(a_1, A) = 1 - \gamma_i(a_1, A)$. For the inductive step, assume that the equality we want to establish holds for $\ell - 1$, i.e., $1 - \sum_{k \le \ell-1} p_i(a_k, A) = \prod_{k \le \ell-1}(1 - \gamma_i(a_k, A))$. Then use equation (4) to get

$$\begin{aligned} 1 - \gamma_i(a_\ell, A) &= 1 - \frac{p_i(a_\ell, A)}{1 - \sum_{k \le \ell-1} p_i(a_k, A)} \\ &= \frac{1 - \sum_{k \le \ell} p_i(a_k, A)}{1 - \sum_{k \le \ell-1} p_i(a_k, A)} \\ &= \frac{1 - \sum_{k \le \ell} p_i(a_k, A)}{\prod_{k \le \ell-1}(1 - \gamma_i(a_k, A))} \\ \implies \prod_{k \le \ell}(1 - \gamma_i(a_k, A)) &= 1 - \sum_{k \le \ell} p_i(a_k, A) \end{aligned}$$

So, this expression holds for ℓ, establishing the claim. Accordingly, equation (4) implies that

$$p_i(a_m, A) = \gamma_i(a_m, A) \prod_{k \le m-1} (1 - \gamma_i(a_k, A)),$$

for all $m \in \{2, ..., M\}$. Hence, the representation holds for all $A \in \mathcal{X}$, and $i \in I$.

Verifying that the clustering is homophilous: We know that $N(i) = R(i)$ by the definition of the clusters. Let N^s and N^t be two distinct clusters, with $i \in N^s$ and $j \in N^t$. Then, $N^s = R(i)$ and $N^t = R(j)$, with $R(i) \neq R(j)$. Since i and j are not connected, there exists a menu A such that $p_i(a, A) = 1$ or $p_j(a, A) = 1$ for some $a \in A$, but not both. Suppose w.l.o.g. that $p_i(a,\, A) = 1$. Then, $p_{i'}(a,\, A) = 1$ for all $i' \in N^s = R(i)$. In establishing the representation, we showed that $p_{i'}(a,\, A) = 1$ implies $a = \max_{\succ_{i'}} A$. We now want to show that $\max_{\succ_{j'}} A \neq a$ for some $j' \in R(j)$.

Choose $j' \in R(j)$ such that $p_{j'}(a,\, A) \leq \mu_j(a,\, A)$. Of course, such a j' must exist. Furthermore, $p_j(a,\, A) \neq 1 \implies \mu_j(a,\, A) < 1$. If $a = \max_{\succ_{j'}} A$, then based on the fact that choice probabilities are according to the representation as shown above,

$$\begin{aligned} p_{j'}(a,\, A) &= \gamma_{j'}(a,\, A) \\ &= \beta_i + (1-\beta_i)\mu_j(a,\, A) \\ &> \mu_j(a,\, A) \end{aligned}$$

However, this is not true by assumption, which implies that $a \neq \max_{\succ_{j'}} A$. Therefore, A is not top-agreeable for $N^t = R(j)$, and the clustering is homophilous.

A.6 Proof of Proposition 3

Let $a, b \in A$ be s.t. $a \succ_i b$. If A is top agreeable for $N(i)$, i.e., for all $j \in N(i)$, the $\succ_j$-best alternative in A is the same, say d, then it follows that for all such j, $p_j(d, A) = 1$. Thus $p_j(c, A) = 0$ for all $c \in A \setminus d$. If $d = a$, then $p_i(a, A) = 1$ and clearly $p_i(a, A) > p_i(b, A)$. If $d \neq a$, then we have $p_i(a, A) = p_i(b, A) = 0$.

Now consider when menu A is not top agreeable for $N(i)$. Then, it follows that $p_i(c, A) \in (0, 1)$, for any $c \in A$. Since $\beta_i \in (0, 1)$,

$$\gamma_i(a, A) = \beta_i + (1-\beta_i)\frac{\sum_{j \in N(i)} p_j(a, A)}{|N(i)|} > \beta_i \geq 0.5$$

and so $\gamma_i(a, A) > 0.5 > 1 - \gamma_i(a, A)$.

Note that $\{c \in A : c \succ_i a\} \subseteq \{c \in A : c \succ_i b\}$. Let

$$M = \prod_{\{c \in A : c \succ_i a\}} \big(1 - \gamma_i(c, A)\big)$$

Then observe that

$$p_i(a, A) = \gamma_i(a, A) M$$

and

$$p_i(b, A) = \gamma_i(b, A)\big(1 - \gamma_i(a, A)\big) M \prod_{\{c \in A : a \succ_i c \succ_i b\}} \big(1 - \gamma_i(c, A)\big)$$

Since $\gamma_i(c, A), \gamma_i(b, A) \in (0, 1)$ and $\gamma_i(a, A) > 0.5$, we have

$$p_i(b, A) \leq \gamma_i(b, A)\big(1 - \gamma_i(a, A)\big) M < \gamma_i(b, A)\gamma_i(a, A) M < \gamma_i(a, A) M = p_i(a, A)$$

That is, $p_i(a, A) > p_i(b, A)$ as required.

A.7 Proof of Proposition 6.2

First we show that $p_i(a^*, ab)$ decreases as β increases. Let $\lambda = \frac{k}{n}$. Taking the first derivative of $p_i(a^*, ab)$ with respect to β:

$$\frac{\partial p_i(a^*, ab)}{\partial \beta} = \frac{2(1-\beta)^2\lambda(1-\lambda) + \sqrt{1-4(1-\beta)^2\lambda(1-\lambda)} - 1}{(1-\beta)^3\lambda(1-\lambda)\sqrt{1-4(1-\beta)^2\lambda(1-\lambda)}}$$

For ease of notation, let $\alpha = 2(1-\beta)^2\lambda(1-\lambda)$. Then rewriting the above equation,

$$\frac{\partial p_i(a^*, ab)}{\partial \beta} = \frac{\alpha - 1 + \sqrt{1-2\alpha}}{(\alpha/2)(1-\beta)\sqrt{1-2\alpha}}$$

It is easy to see that the denominator is positive. Note $\beta \in (0,1)$ and $\lambda(1-\lambda)$ attains a maxima at $\lambda = \frac{1}{2}$, this implies $0 < \alpha < \frac{1}{2}$. It is straightforward to see that $\alpha - 1 + \sqrt{1-2\alpha}$ is decreasing in α for $\alpha \in (0, \frac{1}{2})$. As $\alpha \to 0$, the numerator, $(\alpha - 1 + \sqrt{1-2\alpha}) \to 0$ from above. Thus the numerator is negative and $\frac{\partial p_i(a^*,ab)}{\partial \beta} < 0$.

Second, we show that $p_i(a^*, ab)$ attains a maxima at $\lambda = \frac{1}{2}$. Consider the following expression for the first derivative w.r.t λ:

$$\frac{\partial p_i(a^*, ab)}{\partial \lambda} = \frac{2(2\lambda-1)\left(\sqrt{1-4(1-\beta)^2\lambda(1-\lambda)} - 2(1-\beta)\lambda\left(-2\beta\lambda + \sqrt{1-4(1-\beta)^2\lambda(1-\lambda)} + \beta + 2\lambda - 2\right) - 1\right)}{\lambda^2\sqrt{1-4(1-\beta)^2\lambda(1-\lambda)}\left(2\beta(1-\lambda) + \sqrt{1-4(1-\beta)^2\lambda(1-\lambda)} + 2\lambda - 1\right)^2}$$

Clearly the first derivative has a root at $\lambda = \frac{1}{2}$. It can also be verified that there is no other root. It is straightforward to see that the second derivative at $\lambda = \frac{1}{2}$ given below is negative and thus $p_i(a^*, ab)$ attains a maxima at $\lambda = \frac{1}{2}$.

$$\frac{\partial^2 p_i(a^*, ab)}{\partial \lambda^2}_{|\lambda=\frac{1}{2}} = \frac{8\left(\sqrt{(2-\beta)\beta} - 1\right)}{\sqrt{(2-\beta)\beta} + (2-\beta)\beta}$$

References

Aiello, Luca Maria, Alain Barrat, Rossano Schifanella, Ciro Cattuto, Benjamin Markines, and Filippo Menczer. 2012. "Friendship prediction and homophily in social media." *ACM Transactions on the Web (TWEB)* 6 (2):1–33.

Barber, Michael and Jeremy C Pope. 2019. "Does party trump ideology? Disentangling party and ideology in America." *American Political Science Review* 113 (1):38–54.

Borah, Abhinash and Christopher Kops. 2018. "Choice via social influence." *Unpublished manuscript* .

Brady, Richard L and John Rehbeck. 2016. "Menu-dependent stochastic feasibility." *Econometrica* 84 (3):1203–1223.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of peer effects through social networks." *Journal of econometrics* 150 (1):41–55.

Caplin, Andrew, Mark Dean, and John Leahy. 2018. "Rational inattention, optimal consideration sets, and stochastic choice." *The Review of Economic Studies* 86 (3):1061–1094.

Chambers, Christopher P, Tugce Cuhadaroglu, and Yusufcan Masatlioglu. 2023. "Behavioral influence." *Journal of the European Economic Association* 21 (1):135–166.

Chambers, Christopher P, Yusufcan Masatlioglu, and Christopher Turansick. 2021. "Influence and Correlated Choice." *arXiv preprint arXiv:2103.05084* .

Ciesielski, Krzysztof Chris and Jakub Jasinski. 2016. "On fixed points of locally and pointwise contracting maps." *Topology and its Applications* 204:70–78.

Cohen, Geoffrey L. 2003. "Party over policy: The dominating impact of group influence on political beliefs." *Journal of Personality and Social Psychology* 85 (5):808–822.

Conover, Michael, Jacob Ratkiewicz, Matthew Francisco, Bruno Goncalves, Filippo Menczer, and Alessandro Flammini. 2011. "Political polarization on Twitter." In *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 5. 89–96.

Cuhadaroglu, Tugce. 2017. "Choosing on influence." *Theoretical Economics* 12 (2):477–492.

Dardanoni, Valentino, Paola Manzini, Marco Mariotti, and Christopher J Tyson. 2020. "Inferring cognitive heterogeneity from aggregate choices." *Econometrica* 88 (3):1269–1296.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli. 2010. "Identification of social interactions through partially overlapping peer groups." *American Economic Journal: Applied Economics* 2 (2):241–275.

DellaPosta, Daniel, Yongren Shi, and Michael Macy. 2015. "Why do liberals drink lattes?" *American Journal of Sociology* 120 (5):1473–1511.

Dimant, Eugen. 2024. "Hate trumps love: The impact of political polarization on social preferences." *Management Science* 70 (1):1–31.

Ehlert, Alexander, Martin Kindschi, René Algesheimer, and Heiko Rauhut. 2020. "Human social preferences cluster and spread in the field." *Proceedings of the National Academy of Sciences* 117 (37):22787–22792.

Eliaz, Kfir and Ran Spiegler. 2011. "Consideration sets and competitive marketing." *The Review of Economic Studies* 78 (1):235–262.

Ellis, Wendy E and Lynne Zarbatany. 2017. "Understanding processes of peer clique influence in late childhood and early adolescence." *Child Development Perspectives* 11 (4):227–232.

Ennett, Susan T and Karl E Bauman. 2000. "Adolescent social networks: Friendship cliques, social isolates, and drug use risk." In *In W.B. Hansen, S.M. Giles & M. Fearnow- Kenney (eds.).* Tanglewood Research Estados Unidos, 83–92.

Fershtman, Chaim and Uzi Segal. 2018. "Preferences and social influence." *American Economic Journal: Microeconomics* 10 (3):124–42.

Hall, Jeffrey A and Thomas W Valente. 2007. "Adolescent smoking networks: The effects of influence and selection on future smoking." *Addictive behaviors* 32 (12):3054–3059.

Hefti, Andreas. 2015. "A differentiable characterization of local contractions on Banach spaces." *Fixed Point Theory and Applications* 2015:106.

Jungck, Gerald. 1982. "Local Radial Contractions - A Counter-Example." *Houston Journal of Mathematics* 8:501–506.

Kahan, Dan. 2010. "Fixing the communications failure." *Nature* 463 (7279):296–297.

Kashaev, Nail, Natalia Lazzati, and Ruli Xiao. 2023. "Peer Effects in Consideration and Preferences." *arXiv preprint arXiv:2310.12272* .

Lazzati, Natalia. 2020. "Codiffusion of Technologies in Social Networks." *American Economic Journal: Microeconomics* 12 (4):193–228.

Lin, Xu. 2010. "Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables." *Journal of Labor Economics* 28 (4):825–860.

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.

Lodder, Gerine MA, Ron HJ Scholte, Antonius HN Cillessen, and Matteo Giletta. 2016. "Bully victimization: Selection and influence within adolescent friendship networks and cliques." *Journal of youth and adolescence* 45 (1):132–144.

Macy, Michael, Sebastian Deri, Alexander Ruch, and Natalie Tong. 2019. "Opinion cascades and the unpredictability of partisan polarization." *Science Advances* 5 (8):eaax0754.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." *The review of economic studies* 60 (3):531–542.

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.

Miller, Patrick R and Pamela Johnston Conover. 2015. "Red and blue states of mind: Partisan hostility and voting in the United States." *Political Research Quarterly* 68 (2):225–239.

Paxton, Susan J, Helena K Schutz, Elenor H Wertheim, and Sharry L Muir. 1999. "Friendship clique and peer influences on body image concerns, dietary restraint, extreme weight-loss behaviors, and binge eating in adolescent girls." *Journal of abnormal psychology* 108 (2):255–266.

Seo, Dong-Chul and Yan Huang. 2012. "Systematic review of social network analysis in adolescent cigarette smoking behavior." *Journal of School Health* 82 (1):21–27.

Van Bavel, Jay J, Clara Pretus, Steve Rathje, Philip Pärnamets, Madalina Vlasceanu, and Eric D Knowles. 2024. "The costs of polarizing a pandemic: Antecedents, consequences, and lessons." *Perspectives on Psychological Science* 19 (4):624–639.

Zimmerman, David J. 2003. "Peer effects in academic outcomes: Evidence from a natural experiment." *Review of Economics and statistics* 85 (1):9–23.