Homophily in Peer Groups

Mariagiovanna Baccara[*]   Leeat Yariv[†‡]

October 22, 2012

ABSTRACT.   The focus of this paper is the endogenous formation of peer groups. In our model agents choose peers before making contributions to public projects, and they differ in how much they value one project relative to another. Thus, the group's preference composition affects the type of contributions made. We characterize stable groups and find that they must be sufficiently homogeneous. We also provide conditions for some heterogeneity to persist as the group size grows large. In an application in which the projects entail information collection and sharing within the group, stability requires more similarity among extremists than among moderate individuals.

JEL classification: D82, D85.

Keywords: Homophily, Group Formation, Information Collection.

1. Introduction

There are many realms in which individuals choose *whom* they interact with, socially and strategically. Individuals choose which Internet forums to participate in, clubs to join, neighborhoods to live in, schools to go to, and so on. New platforms such as online social networks, blogs, etc. allow users to choose their peers without any physical or geographical constraint. Interestingly, a vast empirical literature in sociology suggests consistent patterns of group formation. Indeed, individuals exhibit *homophily* – they tend to associate with those similar to them (e.g., in demographics, political opinions, or beliefs).[1]

While, over the decades, the analysis of strategic interactions across domains has received wide attention, theoretically and empirically, the group of players is usually assumed to be determined exogenously.[2] The focus of the current paper is the analysis of an extended game in which, first, agents *choose* their group of peers and, second, a strategic interaction takes place. The goal is to understand how the interplay between the group formation stage and the strategic interaction stage determine the properties of the peer groups that arise in equilibrium.

We study a model in which agents make contributions to two different public projects, or tasks. The tasks can be metaphors for volunteering, freeware development, participation in student associations' activities, and many more. Each agent's taste is characterized by a parameter in $[0, 1]$, proxying for how much she cares about one task relative to the other. For example, depending on their personal circumstances and demographic characteristics, agents may vary in how much they value local parks (important for families and retirees) as opposed to public transportation (important for employed individuals). Similarly, depending on their personal tastes and hobbies, agents might differ in how much they care about the local church's initiatives as opposed to music festivals. We assume that each individual can make a contribution to at most one task. In our model, agents have the possibility of forming groups and *what defines a peer group is that its members benefit from all contributions made towards the tasks.*

For any fixed group of individuals, we characterize the equilibrium task selection, a mapping from the composition of tastes of the agents in the group into the volume of contributions made toward

---

[1] See the literature review for a brief summary of the work on this phenomenon.
[2] See below for a description of several exceptions.

each task. To capture the notion that new technologies allow individuals to connect on platforms that are not bounded by local geography, we step back and consider the group of peers as an *object of choice*. Depending on tastes, and foreseeing the amount and the type of contributions made within each group, each individual prefers certain peer groups to others.

Stable groups are ones that satisfy natural equilibrium constraints in the group-formation stage. That is, *a group is stable if it is optimal for all its members*. Our first main result provides a characterization of stable groups. We show that stable groups of a fixed size are identified by a partition of the taste parameter range $[0, 1]$ into sub-intervals. In particular, a group is stable if and only if there exists an interval in this partition that contains the taste parameters of *all* the group members. This result suggests that if each member has some leverage in choosing her peers, *stability occurs when tastes are sufficiently close*.[3] Intuitively, a group is stable only when all its members agree on the optimal way to allocate the group's contributions across the two tasks, and this occurs when tastes are sufficiently similar.

The growth of online communities and forums, or even the initial introduction of email or sms, allow individuals to connect to one another with greater ease. To address the effects of these innovations on socialization patterns, we look at how stable groups are affected by arbitrarily increasing group size. We provide conditions under which as the group size grows large, stability remains consistent with groups composed of members of *different*, although sufficiently close, tastes–that is, intervals in the partition *do not* converge to singletons.

Our baseline notion of stability is a strong one, in that each individual can potentially deviate to groups composed of agents with *any* taste combination. This framework fits environments in which the population is very large, such as those pertaining to many online communities and social networks. Nonetheless, in smaller populations, deviations are restricted by the existing partition of agents into groups. We study the case of a small population of agents in Section 4. We call a *partition* of the population *stable* if no agent prefers to join a group (an element in the partition) different than the one she is in, or to remain by herself. Certainly, the grand coalition containing all agents is always

---

[3]In fact, new technologies that allow less constrained choices of peers correspond to the emergence of more homogeneous groups. For example, Lynd and Lynd (1929) illustrates how the introduction of the automobile increased the prominence of clubs and coincided with an increase in peer connections based on common interests.

stable and is the most efficient stable allocation. We show that equilibrium sorting is robust to this environment, and in any stable partition similar agents tend to cluster together. Indeed, stable partitions are 'consecutive,' whenever two types belong to the same partition element, so do all types that are ordered between them.

Many social connections (such as Internet forums, information networks, or blogs) have information sharing as a driving force. Moreover, information sharing is one of the motives behind a much wider class of social ties, such as friendships, collaborations, and many others. In these peer groups, individuals share information on topics of interest such as life-decisions regarding parenting and retirement, choices concerning consumption goods for different uses (e.g., food and books), hobbies, and so on. In Section 5, we focus on a special application of our model in which tasks stand for issues over which the group can collect and share information, and agents' utilities differ in the relative weight they put on each issue. For example, children's educational prospects are important for younger individuals, while savings are important for individuals nearing retirement. Similarly, depending on their personal attitudes, individuals may differ in how much they are concerned with the quality of food they eat relative to the selection of books they read. In terms of our application, we note that relevant information can be collected and shared on all of these issues.

In this application, the intervals identifying stable groups exhibit interesting comparative statics. Specifically, these intervals are wider for moderate tastes and become narrower as tastes become more extreme. This implies that *stability requires more similarity for extreme individuals than for moderate ones.* Our last result for the information sharing application relates to a small population and provides conditions under which full segregation, a partition into groups that contain only agents with *the same* taste, can arise as a stable allocation. We exploit the comparative statics developed before to show that, in a similar spirit, *segregation is easier to sustain for individuals of extreme tastes than of moderate ones.*

Related Literature. Lazarsfeld and Merton (1954) coined the term *homophily* – literally meaning "love for the same" – capturing the tendency of socially connected individuals to be similar to one another.[4] In recent years, there has been a growing body of work identifying homophily across fields,

---

[4] The observation that people connect to those similar to them is, in fact, a rather old one. Aristotle remarked in his *Rhetoric* and *Nichomachean Ethics* that people "love those who are like themselves." Plato commented in *Phaedrus* that

ranging from economics (see Benhabib, Bisin, and Jackson (2010)), to political science (see Huckfeldt and Sprague (1995)), to sociology (see McPherson, Smith-Lovin, and Cook (2001)).

In general, similarity of connected individuals on malleable traits (such as political affiliation, education, etc.) can be rooted in one of two processes: (i) selection, or assortative matching, in which similarity begets association, the process modeled in this paper; or (ii) socialization or convergence, in which social ties generate similarity. One way to disentangle these processes entails a study of exogenous characteristics, such as height or race (see Goeree, McConnell, Mitchell, Tromp, and Yariv (2010), Marmaros and Sacerdote (2006), and Mayer and Puller (2008), who identify significant levels of homophily with respect to these attributes). Another approach is to consider longitudinal data, as in Kandel (1978). She studied adolescent friendships and the extent of similarity across dyadic connections regarding four attributes (frequency of current marijuana use, level of educational aspirations, political orientation, and participation in minor delinquency) at several stages of friendship formation and dissolution. Kandel found that observed homophily was the outcome of a significant combination of both types of processes.

On the theoretical side, several recent papers directly address preferences for similarity. Currarini, Jackson, and Pin (2009) *assume* homophilic behavior and study its consequences in a friendship formation model.[5] Peski (2008) derives a preference for similarity endogenously. He assumes certain properties of preferences over friends (complementarities between direct friends and second-degree friends) and the possibility of confusing people who are similar to one another. The necessity to differentiate friends and enemies as much as possible then leads individuals to form friendships with those who are similar.

The underlying idea that the group of players in a strategic situation is, in itself, endogenous motivates some of the work on club formation (see, e.g., Ellickson, Grodal, Scotchmer, and Zame (1999) and Wooders, Cartwright, and Selten (2006)). The basic model of that literature assumes some form of externality across individuals and studies endogenous group formation (often in a general equilibrium setup) in the presence of these externalities. Our approach differs in that externalities in

---

"similarity begets friendship."

[5]Bramoullé et al. (2012) study a related model of friendship formation in which agents of the same 'group' are more likely to meet (and befriend) one another. They characterize the resulting connection characteristics.

our setting arise only through the public good contributions (specifically, no goods are traded after groups are formed). Furthermore, we focus on the characteristics of the emergent groups (namely, the distribution of tastes as a function of the environment's fundamentals).

Several elements of our model are reminiscent of work in other areas. First, the idea that agents may choose peer groups that match their preferences is an ongoing theme in the theory of public choice, going back to Tiebout (1956). These models define municipalities by the government services and tax rates that they offer. Individuals choose a community that maximizes their utility. Nonetheless, the strategic interaction that follows and the structure of utilities are very different. Furthermore, much of that work is concerned with the efficiency of such processes, rather than with the similarity or heterogeneity of equilibrium communities.[6] Second, the notion that agents optimally select those with whom they communicate appears also in Calvó-Armengol, De Martí, and Prat (2011). They consider a set of connected agents who differ in the accuracy of their *exogenously* provided private information, and are ex-ante identical otherwise. In contrast, we characterize the endogenous similarity within groups in which all agents freely communicate with all others.

Recently, there has been a proliferation of work illustrating the potential explanatory power of social connections regarding individual outcomes across contexts, covering public goods provision, crime, job search, political alliances, trade, friendships, and information collection.[7] Particularly in view of the vast literature on homophily, an important empirical issue in this literature is that correlations between behavior and outcomes of individuals and their peers may be driven by common unobservables and, therefore, be spurious (see Evans, Oates, and Schwab (1992) and Manski (1993, 2000)). Understanding similarity patterns is potentially important for mitigating such endogeneity problems.

---

[6]Regarding this line of research, our results in Section 4 share some common elements with Greenberg and Weber (1986) and Demange (1994). We refer to that section for a comparison between our approach and theirs.

[7]This literature is too extensive to survey here. Important work includes Coleman (1966), Conley and Udry (2010), Foster and Rosenzweig (1995), Glaeser, Sacerdote, Scheinkman (1996), Granovetter (1994), Katz and Lazarsfeld (1954), and Topa (2001).

2. The Model

There is a population of agents. Each agent is characterized by a taste parameter $t \in [0, 1]$. The agents are divided into groups of size $n \geq 2$.[8] A group may consist of agents with different parameters. Each agent takes one of two actions, A and B. The payoff of an agent $U(t, k_A, k_B)$ depends on his own taste parameter t, and on the number of actions A or B taken by all the members of the group, respectively k_A and $k_B = n - k_A$. We interpret A and B as two tasks to which agents can contribute. Tasks can stand for different public projects, different issues on which to collect information, etc. Each agent can select simultaneously which task to make one contribution to.[9] Throughout the paper we assume that contributions are free. This may fit environments in which a contribution does not entail a substantial effort (say, reading different parts of the newspaper), or situations in which some form of contribution is par for the course, and individuals face only a choice of what task to focus on– e.g., which websites to visit while surfing the Internet (those directed at child-rearing or those focusing on investment advice), which volunteer groups to join, which part of the newspaper to read in the morning (the food section or the book reviews), and so on.

For the sake of technical convenience, we will assume that $U(t, k_A, k_B)$ is defined over $[0, 1] \times \mathbb{R} \times \mathbb{R}$ and is twice continuously differentiable with respect to all arguments, increasing in k_A and k_B, and that $\frac{\partial^2 U}{\partial k_A^2} < 0, \frac{\partial^2 U}{\partial k_B^2} < 0, \frac{\partial^2 U}{\partial k_A \partial k_B} \geq 0, \frac{\partial^2 U}{\partial t \partial k_A} > 0$, and $\frac{\partial^2 U}{\partial t \partial k_B} < 0$.[10] In words, outcomes improve with additional contributions, but exhibit decreasing marginal returns to contributions on either task. Furthermore, an increase in contributions toward one task does not reduce the marginal returns from contributions to the other task. Last, types are a proxy for how much agents care about task A relative to task B, with higher parameter values of t corresponding to agents who care more about contributions directed at task A. In particular, these assumptions imply that $\frac{\partial^2 U(t, k, n-k)}{\partial t \partial k} > 0$ and $U(t, k, n - k)$ is strictly

---

[8] Assuming that the group size is exogenous captures situations in which agents face a fixed contraint on how much time and effort they can invest in communication, or social interactions. While the group size is exogenous throughout Section 3, it becomes endogenous in the analysis in Section 4. We do note that some evidence suggests that there is a cognitive cap on the number of active social ties humans can maintain. Projecting from primates, Dunbar (1992) estimated this number to be around 150, while more recent work suggests that the number in modern times is closer to 300 (see, e.g., McCarty, Killworth, Bernard, Johnsen, and Shelley (2000)).

[9] The analysis of the paper does not change if we assume that each agent can make any fixed number of contributions $h \geq 1$.

[10] While the number of agents contributing to either task is, by definition, discrete, the assumption that the utility is defined over continuous variables allows us to use calculus and makes the presentation of assumptions and results easier. Had we defined utility over integer arguments for the task contributions, we could choose a smoothing of the function that would correspond to the utility functions considered here.

concave in k.[11]

A large class of utility functions satisfies the above assumptions. For instance, consider two production functions f_A and f_B, increasing and concave, that map effort into utilities derived from either task. Assume that the general utility for an agent of taste $t \in [0,1]$, who is in a group of n agents, k_A of whom invest their unit of effort in dimension A and $k_B = n - k_A$ of whom invest their unit of effort in dimension B, is given by the weighted average:

$$U(t, k_A, k_B) \equiv t f_A(k_A) + (1-t) f_B(k_B),$$

which satisfies all assumptions above.[12] Another example of utility that satisfies the assumptions above is the Cobb-Douglas utility $U(t, k_A, k_B) \equiv (k_A)^t (k_B)^{1-t}$.[13]

For any group composed of agents with tastes $(t_1, ..., t_n)$, expected payoffs are ultimately identified by the profile of chosen tasks $(x_1, ..., x_n)$, where $x_i \in \{A, B\}$ is the task chosen by agent i. We call the induced game the *task–selection game*. As a *tie-breaking rule*, we assume that an agent who is indifferent between tasks A and B makes an A-contribution (this simplifies the exposition, but is not crucial for our analysis). We focus on equilibria in pure strategies. As it turns out, given our tie-breaking rule, a pure equilibrium exists and the volume of equilibrium A- and B-contributions is determined uniquely, as the following lemma guarantees.

Lemma 1 (Existence and Uniqueness) *For any group of n agents with tastes $t_1 \geq t_2 \geq ... \geq t_n$, there exists $k^* \in \{0, ..., n\}$ such that all agents $i \leq k^*$ selecting task A, and all agents $i > k^*$ selecting task B, is part of a Nash equilibrium of the task-selection game. Furthermore, all Nash equilibria of the task-selection game entail the same number k^* of agents selecting task A.*

---

[11]Indeed,

$$\frac{\partial^2 U(t, k, n-k)}{\partial t \partial k} = \frac{\partial^2 U(t, k, n-k)}{\partial t \partial k_A} - \frac{\partial^2 U(t, k, n-k)}{\partial t \partial k_B} > 0.$$

Furthermore,

$$\frac{d^2 U(t, k, n-k)}{dk^2} = \frac{\partial^2 U(t, k, n-k)}{\partial k_A^2} + \frac{\partial^2 U(t, k, n-k)}{\partial k_B^2} - 2 \frac{\partial^2 U(t, k, n-k)}{\partial k_A \partial k_B} < 0.$$

[12]In Section 5, we focus on a particular incidence of this family, in which the production functions are $f_A(k^A) = 1 - \frac{1}{2}(1 - q_A)^{k^A}$ and $f_B(k^B) = 1 - \frac{1}{2}(1 - q_B)^{k^B}$, corresponding to an environment in which contribution to a task translates into the collection of a potentially informative signal.

[13]This example requires us to restrict the definitions above to $k_A, k_B \in \mathbb{R}_+$ or, alternatively, to extend the function to $[0,1] \times \mathbb{R} \times \mathbb{R}$.

3. Stable Groups in Infinite Populations

In this section we consider an extended game composed of two stages. First, for a given group size n, each agent of taste $t \in [0,1]$ can choose the tastes of the remaining $n-1$ agents in her group.[14] Second, the task-selection game described above is played. Since Lemma 1 guarantees that the equilibrium volume of A-contributions is determined uniquely in the task-selection game, the agent's optimization problem in the first stage of this extended game is well defined. We denote the set of optimal groups chosen by agent t at the first stage by $O(t)$, each element of which contains t as a member. We define stability in the first stage of the extended game as follows.

Definition (Stable Group) *A group* $(t_1, .., t_n)$ *is stable if it is optimal for all its members – i.e.,*
$$(t_1, .., t_n) \in \cap_{i=1}^n O(t_i).$$

In a stable group, each agent maximizes her expected utility given the tastes of others in the group, foreseeing the equilibrium played in the task-selection game that ensues. This notion of stability is, therefore, a natural equilibrium condition for the group-selection stage in an environment that allows individuals to connect in an unconstrained way (except for the fixed group size n).

The goal of this section is to analyze the group properties entailed by this stability notion. First, denote by $n^A(t)$ the optimal number of A-contributions an agent with taste parameter t would choose out of a total of n available contributions. That is, given $t \in [0,1]$, $n^A(t)$ is the maximal integer k such that

$$U(t, k, n-k) \geq U(t, k-1, n-k+1) \tag{1}$$

is satisfied. If (1) is not satisfied for any k, we define $n^A(t) = 0$. Let $n^B(t) \equiv n - n^A(t)$. Therefore, $(n^A(t), n^B(t))$ represents the **unconstrained optimal allocation** of n contributions for an agent of taste t. Our assumptions on the underlying utility functions together with the tie-breaking rule assure that $U(t, k, n-k)$ is strictly concave in k, and the definition of $n^A(t)$ indeed corresponds to the (generically) unique optimal number of A-contributions for an agent of type t. As t increases, an agent with taste parameter t cares more about task A and so that agent's optimal allocation entails

---

[14] For now, we assume that the pool of potential agents to choose from is infinite and that any combination of tastes is feasible. In Section 4, we restrict the set of agents to a finite population.

more contributions directed at that task. In particular, $n^A(t)$ increases with t.[15] Furthermore, our assumptions guarantee that $n^A(t)$ increases with the group size n.

Given a group size n, consider the optimal group composition from the point of view of an agent with taste parameter $t \in [0,1]$. Any first-best group for an agent with taste parameter t is composed so that $n^A(t)$ agents make an A-contribution, and $n^B(t)$ agents make a B-contribution, therefore achieving the unconstrained optimal allocation for an agent of taste t. Groups consisting of all agents sharing the taste parameter t are, therefore, optimal. Nonetheless, an optimal group for the agent of taste t can also be composed of just the right number of extremists on each side, thereby achieving maximal polarization.[16]

Since an agent's optimal group entails her unconstrained optimal allocation of n contributions across the tasks A and B in the task-selection game, in a *stable* group in which all agents optimize on their peers' tastes, all agents have to agree on the optimal allocation of contributions across the two tasks. In particular, a group formed by identical agents is always stable and *stable groups always exist* in this setting. More generally, stable groups are characterized as follows.

Proposition 1 (Equilibrium Sorting) *For any group size n, there exists v and v', with $0 \leq \nu \leq v' \leq n$ such that, letting $T^n_k \equiv \{t \mid n^A(t) = k\}$ for $v' \leq k \leq v$, then:*

(a) The sets $\{T^n_k\}^{\nu'}_{k=v}$ are a sorted partition of the interval $[0,1]$, where $T^n_v = [0, t(v+1))$, $T^n_k = [t(k), t(k+1))$, for $k = v, ..., \nu' - 1$, and $T^n_{\nu'} = [t(\nu'), 1]$;

(b) A group comprised of agents with tastes $(t_1, ..., t_n)$ is stable if and only if there exists $k = v, ..., \nu'$, such that for all $i, t_i \in T^n_k$. That is, all taste parameters in the group belong to the same element of the partition.

Proposition 1 guarantees that stable groups can be formed only by agents whose tastes are close enough – namely, they lie in one of the intervals T^n_k. Intuitively, consider the optimal number of A-contributions $n^A(t)$ that each agent of taste t would choose. Recall that, since types are a proxy for how much agents care about task A relative to task B, $n^A(t)$ is an increasing function. Each

---

[15] See the proof of Proposition 1 in the Appendix for details.

[16] See Step 1 of the proof of Proposition 1 for a characterization of the set of optimal groups, $O(t)$, for any agent of taste t.
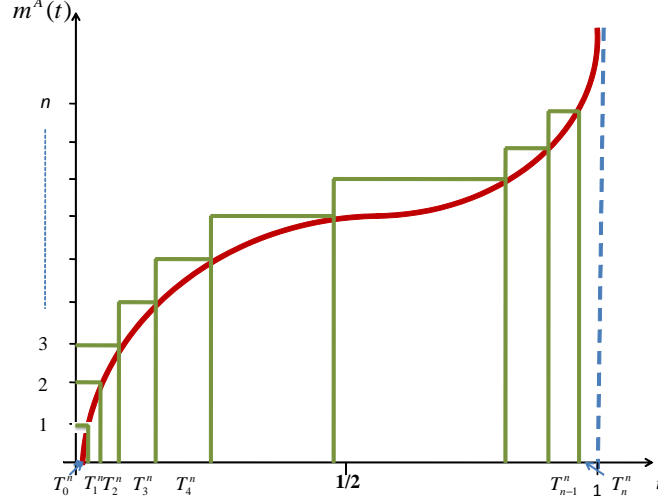
Figure 1: Stable Groups

interval in the partition $\{T_k^n\}_k$ includes all taste parameters of agents who agree on a given optimal allocation of contributions, namely all agents for whom $n^A(t) = k$ for some k. The monotonicity of $n^A(t)$ guarantees that the set of types satisfying such an equality is an interval.

In order to easily depict this result, and describe some of the results that follow, it is useful to consider a smoothing of $n^A(t)$. Specifically, whenever there is a real number achieving equality within constraint (1), under our assumptions on U that number is unique, and we denote it by $m^A(t)$; If the left (or right) hand side of constraint (1) is always greater than the right (or left) hand side, we denote $m^A(t) = n$ (or $m^A(t) = 0$). In what follows, we will assume that $m^A(t)$ is differentiable over $(0,1)$.[17] We get that $n^A(t) = \lfloor m^A(t) \rfloor$ if $m^A(t) \in [0,n]$, $n^A(t) = 0$ if $m^A(t) < 0$, and $n^A(t) = n$ if $m^A(t) > n$. Roughly speaking, $m^A(t)$ captures the point at which the agent is indifferent between staying with her current allocation and shifting one contribution toward the B-task. The function $m^A(t)$ is depicted in Figure 1 for a case in which the number of intervals is maximal, i.e., $\nu = 0$ and $v' = n$. Each interval in the partition $\{T_k^n\}_k$ includes all taste parameters of agents who agree on a given optimal allocation of contributions. In Figure 1, agreement on the number of A-contributions $n^A(t) = k$ corresponds to

---

[17] As mentioned before, this assumption is not restrictive in that for any U defined on discrete contributions, we could find an extension of it over the real line that satisfies this constraint. Notice further that whenever there is a solution to $U(t, m^A(t), n - m^A(t)) = U(t, m^A(t) - 1, n - m^A(t) + 1)$, it is unique due to our assumptions on the decreasing marginal returns of contributions on each task.

the interval of taste parameters that is projected from $[k, k+1)$ on the y-axis. That is, the interval of tastes for which $m^A(t) \in [k, k+1)$ and $n^A(t) = k$.

3.1. Group Size. Many recent technologies, such as email, instant messaging, online networks, etc., allow individuals to connect to one another with greater ease. It has been empirically observed that larger groups tend to be characterized by an increased degree of similarity.[18] In light of this evidence, we now look at how stable groups are affected by arbitrarily increasing group size.[19]

Consider first the Cobb-Douglas example, where $U(t, k_A, k_B) \equiv (k_A)^t (k_B)^{1-t}$. It is straightforward to see that the optimal number of A-contributions for an individual of type t is given by $n^A(t) \in \{\lfloor nt \rfloor, \lceil nt \rceil\}$. In particular, as n grows, the intervals of agent types that agree on the allocation of contributions shrink at a rate of $1/n$. That is, stable groups become asymptotically homogeneous. As it turns out, this is not always the case.

As before, we denote by $\{T_k^n\}_{k=0}^{\nu}$ the partition of the unit interval into sub-intervals defining stability. Since each stable interval is a projection of the function $n^A(t)$ on the x axis, the 'flatter' is that function, the larger are the projected intervals. It is therefore intuitive to presume that the behavior of the intervals $\{T_k^n\}_{k=0}^{\nu}$ as n diverges is linked to the behavior of the slope of $m^A(t)$, the continuous approximation of $n^A(t)$, as suggested by Figure 1.

As it turns out, the intervals in which the slope of $m^A(t)$ is bounded above as n increases remain of substantial size and do not converge to singletons. This guarantees that some degree of heterogeneity persists in stable groups as n diverges.

Proposition 2 (Large n) *Suppose there exists $[a, b] \subset [0, 1]$ in which $\frac{dm^A(t)}{dt}$ is uniformly bounded for all n. Then, there exists L such that $|\{k \mid T_k^n \cap [a, b] \neq \varnothing\}| \leq L$ for all n.*

Intuitively, from Figure 1, whenever the slope of $m^A(t)$ is bounded by M over an interval $[a, b]$, the image of $[a, b]$ corresponds to an interval on the y-axis of size at most $M(b-a)$. Now, stable groups are determined according to the projection of intervals between integers on the y-axis via $m^A(t)$ on the type-axis. Therefore, the number of stable groups agents with taste parameters in $[a, b]$ can belong to

---

[18]See, for instance, Currarini, Jackson, and Pin (2009) and references therein.

[19]Note that the comparative statics with respect to group size n are fully isomorphic to a comparative statics exercise in which the group size is fixed, but we increase the number of contributions that each agent makes from 1 to any $h \geq 1$. In this case, group stability requires all agents in the group to agree on how to allocate $n \times h$ contributions.

is at most the number of such intervals (those between integers on the y-axis) that $m^A([a,b])$ intersects, which is bounded.

In terms of fundamentals, using the definition of $m^A(t)$ and the Implicit Function Theorem, we get

$$\frac{dm^A(t)}{dt} = \frac{\frac{\partial U(t,m^A,n-m^A)}{\partial t} - \frac{\partial U(t,m^A-1,n-m^A+1)}{\partial t}}{-\frac{\partial U(t,m^A,n-m^A)}{\partial k_A} + \frac{\partial U(t,m^A,n-m^A)}{\partial k_B} + \frac{\partial U(t,m^A-1,n-m^A+1)}{\partial k_A} - \frac{\partial U(t,m^A-1,n-m^A+1)}{\partial k_B}}. \qquad (2)$$

From (2), it is immediate to translate the condition of Proposition 3 into conditions on the underlying utilities. For instance, whenever the two tasks are independent, so that $\frac{\partial^2 U(t,k_A,k_B)}{\partial k_A \partial k_B} = 0$, and all other second derivatives are uniformly bounded away from 0 for all n, the condition is satisfied.[20] The information sharing application presented in Section 5 represents one example in which the condition is satisfied.

4. STABILITY IN A FINITE POPULATION

Thus far, our notion of stability has imposed no restrictions on the groups available for the agents to join. Indeed, agents contemplate all possible combinations of tastes when choosing their optimal peer group. This is a good description for very large (strictly speaking, infinite) populations, and allows us to derive a clean characterization of stable groups in different contexts. However, when a finite population of agents is partitioned into groups, there is a restricted set of groups that is conceivably available to an agent. In this section, we study partitions of agents into groups that are "endogenously" stable. That is, we look for partitions of the population into groups such that no feasible deviation of an agent to a different existing group (or to a singleton), is profitable.

Suppose, then, that there is a finite set of agents $N = \{1,...,l\}$. Let $T = \{t_1,..,t_r\}$ be the set of tastes represented in N, and denote by $N_i \subseteq N$ the set of agents with taste t_i, $|N_i| = m_i$ for $i = 1,..,r$. Thus, $N = \cup_{i=1}^r N_i$ and, without loss of generality, we assume $t_1 > ... > t_r$. In analogy with our

---

[20] When tasks are independent, from the Mean-Value Theorem, (2) translates into

$$\frac{dm^A(t)}{dt} = \frac{\frac{\partial^2 U}{\partial t \partial k_B} - \frac{\partial^2 U}{\partial t \partial k_A}}{\frac{\partial^2 U}{\partial k_A^2} + \frac{\partial^2 U}{\partial k_B^2}},$$

where derivatives are evaluated at numbers between 0 and n. When the nominator is bounded, and the denominator is bounded away from 0, $\frac{dm^A}{dt}$ is bounded as well. In particular, the condition of Proposition 3 ties to the asymptotic curvature of the underlying utility.

baseline setup, the extended game that the agents in the set N play consists of two stages. First, the population N is partitioned into groups. Let $\mathcal{G} = \{G_1, ..., G_s\}$ be the resulting partition of N.[21] Importantly, here we *do not* exogenously fix the size of the groups composing the partition $\mathcal{G}$. The second stage of the game coincides with the task-selection game described in Section 2.

Within each group G_i, a task-selection equilibrium corresponds to the description in Lemma 1. To complete the analysis of this game, we can focus on the group-formation stage. We define stability in this setting as follows.

Definition (Stable Partition) *A partition* $\mathcal{G} = \{G_1, ..G_s\}$ *is* stable *if there exist no* G_i, $G_j \in \mathcal{G}$ *and* $a \in G_i$ *such that agent* a *prefers either the group* $G_j \cup \{a\}$ *or the singleton* $\{a\}$ *to* G_i.

This notion of a stable partition is reminiscent of the notion of the core, which also requires a type of group stability. Nonetheless, there are several important distinctions. First, the setup is different – cooperative games normally specify group values, rather than individual values *within* groups that are *derived endogenously from a strategic interaction*. Second, cooperative solutions (e.g., the core) are more restrictive in that they allow for arbitrary group deviations, not *only unilateral* ones. Thus, if we take a group's value to be the sum of its members' expected utilities, the set of stable groups we look at corresponds to a superset of the core, which in this setting corresponds to the grand coalition only. Finally, the core identifies stable allocations of resources, rather than a characterization of the emerging partitions themselves. Since adding a member to a group is costless, and a new group member provides more contributions in the task-selection stage, groups always benefit from adding more members. Thus, we do not have to attend to any issues pertaining to the willingness of a group to accept a new member.[22]

Certainly, the *grand coalition* $\mathcal{G} = \{N\}$ is always a stable partition and, in fact, it is the welfare-maximizing partition.[23] In what follows, we study the stability of other partitions. In particular, Proposition 3 addresses two properties of any stable partition. Namely, agents of similar tastes cluster

---

[21] So that $G_i \cap G_j = \varnothing$ for all $i \neq j$ and $\bigcup\limits_{i=1}^{s} G_i = N$.

[22] See Bogomolnaia et al. (2007) for a study of environments in which stability is affected by the willingness of group members to accept an individual migrating from another group to their own.

[23] This is in contrast with the underlying assumption in Demange (1994), who studies a similar question in a setup where the grand coalition generates inefficient outcomes and, unlike here, any collection of agents can jointly deviate.

together, and all agents sharing the same taste must be contained in the same group.

Proposition 3 (Consecutive and Minimal Groups) *In any stable partition* $\mathcal{G}$,

1. *All groups are consecutive, i.e., if $t_i > t_j > t_h$, and two agents with tastes t_i and t_h are in a group $G \in \mathcal{G}$ then any agent with taste t_j must be in G as well;*

2. *All agents with the same taste are contained in the same group, i.e., for each t_i, there exists a unique $G \in \mathcal{G}$ such that $N_i \subseteq G$.*

Point (1) of Proposition 3 is a sorting result in the spirit of Proposition 1: In any stable partition groups are characterized by individuals that are similar enough in taste.[24] Point (2) of Proposition 3 implies that the maximal number of groups contained in a stable partition is bounded by the total number r of different taste parameters in the population. Intuitively, consider the first part of the Proposition. If the agent of type t_j prefers a different group G' than the group G containing agents of types t_i and t_h, then G must involve the collection of either more contributions to task A or more contributions to task B relative to G' (or else the t_j-type agent would benefit by switching to G). Suppose more contributions are directed at task A in G. Since the agent of type t_h cares even more than the t_j-type agent about task A and her switch to G' would assure an even greater overall contribution volume (in G' augmented with her participation), the t_h-type agent cannot be optimizing.

The intuition for point (2) of Proposition 3 is similar. Suppose that two agents a and a' with the same taste parameter t_i belong to two different groups G and G', respectively. Consider the agent a of taste parameter t_i in G. Since $\mathcal{G}$ is stable, this agent must prefer to stay in G rather than being in $G' \cup \{a\}$. However, since a and a' have the same tastes, and a shift of agent a' to group G would entail even more contributions made, this implies that agent a' must prefer being in $G \cup \{a'\}$ rather than in G', which contradicts the stability of $\mathcal{G}$.

5. INFORMATION SHARING IN GROUPS

We now consider a particular application of our model to environments in which contributions provide access to information. Specifically, suppose there are two issues at stake: α and β, each taking a value

---

[24] The notion of 'consecutive groups' is reminiscent of the one adopted by Greenberg and Weber (1986) in the context of characterizing the core in Tiebout multi-jurisdictional economies.

in $\{0, 1\}$. The values of α and β are determined independently at the outset of the game. We assume that each issue $I \in \{\alpha, \beta\}$ has equal probability of receiving the value 0 or 1.[25] Issues can stand for many problems, ranging from choices of the best food shop and bookstore, to selecting physicians in two different areas of expertise (say, a dentist and a pediatrician).

The agent's goal is to match her actions with the realized issues. The taste parameter t measures how much an agent's utility is affected by making the right decision on each issue.[26] The utility the agent derives is the sum of two components. If she makes the right decision on issue α, she receives a payoff of t (and 0 otherwise); If she makes the right decision on issue β, she receives a payoff of $1 - t$ (and 0 otherwise). For example, all agents benefit by choosing a superior supermarket and a superior bookstore, but, depending on their consumption patterns, they may differ in how much one affects their utility with respect to the other. Similarly, agents may be affected differently by the selection of an able dentist relative to a pediatrician depending on their age, health, and family status.

Contributing to a task in this setting translates into the collection of information. In other words, prior to making a decision, each agent can select simultaneously one of two information sources, A or B, corresponding to the two issues. Information source A (a contribution to task A) provides the realized issue α with probability $q_A > 0$. That is, upon choosing information source A, the agent observes a signal $s \in \{0, 1, \varnothing\}$ according to:

$$\Pr(s = \alpha) = q_A, \quad \Pr(s = \varnothing) = 1 - q_A.$$

Similarly, information source B (a contribution to task B) provides the realized issue β with probability $q_B > 0$.

If k^A contributions are made toward task A (i.e., k^A A-signals are collected), the probability of making the right decision on that issue is given by $\left[1 - \frac{1}{2}(1 - q_A)^{k^A}\right]$. Similarly for issue B. The resulting utility for an individual in a group in which k^A contributions are made toward task A and

---

[25] The entire analysis can be extended directly to asymmetric priors.

[26] While in this setting the issues are common value (so, the right decision for all agents is identical), our analysis would follow the same lines if agents had different opinions on what is the right decision conditional on the realized issue.

k^B contributions are made toward task B is given by:

$$U(t, k^A, k^B) \equiv t \left[1 - \frac{1}{2} (1 - q_A)^{k^A} \right] + (1 - t) \left[1 - \frac{1}{2} (1 - q_B)^{k^B} \right], \qquad (3)$$

which satisfies all of our assumptions.

5.1. Stability in an Infinite Population.   The stability concept for the infinite population case introduced in Section 2 is a natural equilibrium condition in an environment that allows individuals to connect in an unconstrained way such as online forums, information networks, blogs, etc.

An agent's optimal group entails her unconstrained optimal allocation of n signals across the sources A and B being selected in the task-selection, or information-collection, game. Thus, in a *stable* group in which all agents optimize on their peers' tastes, all agents have to agree on the optimal allocation of signals across the two sources. In this setting, extreme agents (with taste parameters $t = 0$ or $t = 1$) prefer all signals to be collected on the issue they care about. Proposition 1 then implies that for any group size n, there exists a partition $\{T_k^n\}_{k=0}^n$ of the interval $[0, 1]$, where $T_0^n = [0, t(1))$, $T_k^n = [t(k), t(k+1))$, for $k = 1, ..., n-1$, and $T_n^n = [t(n), 1]$ such that a group comprised of agents with tastes $(t_1, ..., t_n)$ is stable if and only if all taste parameters in the group belong to the same element of the partition. That is, the number of intervals characterizing stable groups is maximal. In the information-sharing application, the following Proposition further describes the partition $\{T_k^n\}_{k=1}^{n-1}$.

Proposition 4 (Information Sharing– Stable Groups) *In the information-sharing case, the length of the intervals $\{T_k^n\}_{k=1}^{n-1}$ is increasing for $k = 1, ..\widehat{k}$ and decreasing for $k = \widehat{k}, .., n-1$, where $\widehat{k}$ is such that $\frac{1}{2} \in T_{\widehat{k}}^n$ or $\frac{1}{2} \in T_{\widehat{k}+1}^n$. Thus, the intervals $\{T_k^n\}_{k=1}^{n-1}$ are narrower for extreme tastes, and wider for moderate tastes.*[27]

The length of the intervals $\{T_k^n\}_{k=0}^n$ provides a proxy for equilibrium homophily: the narrower an interval is, the closer the agents' tastes have to be in order for them to form a stable group. Proposition 4 addresses how the intervals' lengths are affected by the proximity of the intervals to the extreme tastes. In particular, the proposition implies that *stability requires more similarity for*

---

[27] In the Appendix, we show that, under mild conditions on q_α, q_β, and n, this result extends to the entire sequence $\{T_k^n\}_{k=0}^n$. For example, $(1 - q_\alpha)(1 - q_\beta) < 1/2$ and n high enough, are sufficient to guarantee this.

extreme individuals than for moderate ones. We note that the comparative statics of Proposition 4 is tied to the particular notion of distance implied by the linearity of the utility function we consider. From an empirical perspective, we view this specification as a natural one to consider as it would be the one directly obtained from a linear estimation of the utility as a function of the different issues. After a normalization with respect to overall preference intensity, the coefficient obtained in such an estimation would correspond to our parameter t.

In order to see the intuition underlying this result, consider the function $m_A(t)$, which approximates the optimal number of A-signals for an agent of taste t. Observe that by raising t above $\frac{1}{2}$, the relative value of an A-signal increases since the agent cares more about issue α than issue β. However, a countervailing force is in effect. Indeed, when the optimal number of A-signals increases with t (and the optimal number of B-signals decreases), the marginal return of an A-signal relative to a B-signal decreases. For the sake of simple intuition, let us focus on the symmetric case $q_A = q_B \equiv q$. From the definition of $m_A(t)$ and equations (1) and (3), after some manipulations, we get that in this environment $m_A(t)$ satisfies

$$\frac{t}{1-t} = \frac{q_B(1-q_B)^{n-m^A(t)}}{q_A(1-q_A)^{m^A(t)-1}} = (1-q)^{n-2m^A(t)+1}. \tag{4}$$

Both sides of condition (4) are increasing in t, the left hand side capturing the first, direct effect of an increase in t, and the right hand side representing the marginal return of a B-signal relative to an A-signal. Defining the function corresponding to the right hand side of condition (4) as $M(k) \equiv (1-q)^{n-2k+1}$, it is easy to see that $\frac{M(k+1)}{M(k)}$ is equal to a constant, namely, $(1-q)^{-2}$. Thus, the right hand side of condition (4) increases at a constant rate in $m^A(t)$. On the other hand, the left hand side of (4) increases at increasing rates in t when $t > 1/2$. This implies that, as t increases, $m^A(t)$ has to increase at increasing rates to satisfy (4). In other words, $m^A(t)$ must be convex for $t > 1/2$ and concave for $t < 1/2$. This induced shape of $m^A(t)$ implies the property described in Proposition 4 directly.[28,29]

---

[28] The reason why the extreme intervals can generally follow a different pattern is the following. The extreme intervals collect all the taste parameters t such that the problem of the optimal allocation of signals across the two sources has a corner solution. Then, for example, for a given q_B, if both q_A and n are very low, there is a wide range of t's for which it is optimal to get n B-signals. In order for the pattern to carry through for the extreme intervals, the signal accuracies and the group size need to be sufficiently large.

[29] These intuitions can be easily generalized. While here we assume a particular signal generation process, when the

From an empirical point of view, these results are important for understanding the link between group composition and extremism of opinions. Some of the recent social psychology literature (see Myers (2007)) suggests that more homogeneous groups tend to exhibit far more extreme opinions than heterogeneous ones following group interactions. While the social psychology literature focuses on mechanisms by which group dynamics generates extremism, our results imply that it is important to account for the way these groups are created to begin with. Specifically, the comparative statics we identify indicate that more extreme individuals would be more likely to form homogeneous groups in the first place.

As we did in Section 3.1 in the general case, we now look at how stable groups are affected by arbitrarily increasing group size in the information sharing context. As it turns out, we can use the results of Proposition 2 directly. In fact, in the information sharing environment, $\frac{dm^A(t)}{dt}$ is bounded above 0 for any interval $[a, b] \subset (0, 1)$.[30] For agents with extreme types, the optimal contribution to a task is close to the number of group participants n and so, as n increases, the function $m^A(t)$ defined above asymptotes at $t = 0$ and $t = 1$. We call the stable groups that choose all signals from the same source (i.e., have taste parameters in either T_0^n or in T_n^n) *extreme stable groups*. We call all other stable groups *non-extreme stable groups*. Proposition 2 implies that non-extreme stable groups do not converge to singletons. This is not the case for extreme stable groups. These groups contain agents for whom the optimal allocation of n signals is a corner solution, which does not equalize the marginal utility of signals. As the number of signals n increases, more agents whose tastes are not at the extremes of the interval $[0, 1]$ tend to reach interior solutions. We therefore get the following corollary.[31]

---

marginal return of a signal does not decrease "too quickly," the comparative statics described in Proposition 4 hold.

[30] Indeed, in this setting,

$$\frac{dm^A(t)}{dt} = -\frac{1}{t(1-t)\ln\left[(1-q_\alpha)(1-q_\beta)\right]},$$

which implies that $m^A(t)$ shifts upward in a parallel way as n grows large.

[31] We include a formal proof of Part 2 of Corollary 1 in the Appendix.

Corollary 1 (Information Sharing - Large n)

1. *For any $[a, b] \subset (0, 1)$, there exists L such that $|\{k \mid T_k^n \cap [a, b] \neq \varnothing\}| \leq L$ for all n;*

2. *If two agents both belong to an extreme stable group of size n, then they both belong to an extreme stable group of any smaller size $n' < n$.*

It is interesting to compare the results of the corollary with the empirical observations suggesting more homophily in larger groups cited in Section 3.1. Since we indeed find that, as n increases, the extreme intervals tend to break down into an increasing number of smaller intervals, our result does not necessarily contradict this evidence. Instead, it qualifies it. The corollary highlights the fact that the location within the taste spectrum may play an important role in identifying these sort of comparative statics: *The tendency of larger groups to display more similarity should be stronger for extreme taste parameters than for moderate ones.*

5.2. Information Sharing In a Finite Population. We now move to the stability concept associated with a finite population introduced in Section 4. As before, we assume there is a finite set of agents $N = \{1, ..., l\}$. Let $T = \{t_1, ..., t_r\}$ be the set of tastes represented in N, and denote by $N_i \subseteq N$ the set of agents with taste t_i, $|N_i| = m_i$ for $i = 1, ..., r$.

We define the *fully segregated partition* to be the partition in which each set is formed only by agents of the same type, the partition $\mathcal{G} = \{N_1, N_2, ..., N_r\}$. Since Proposition 3 guarantees that agents of the same type cannot be divided across different groups in any stable partition, we can conclude that the fully segregated partition is the least efficient partition that could conceivably be stable. In order for the segregated partition to be stable, two types of restrictions need to hold. First, the different types need to be sufficiently dispersed, so that any agent a deviating to a homogeneous group composed of different-type individuals would find the allocation of contributions in the group far from her optimal allocation. Second, the number of agents of each type needs to be comparable: if $m_i >> m_j$, agents of type t_j would have stronger incentives to join N_i in order to benefit from the volume of contributions made in that group. To focus on the first issue (identifying taste distributions that allow for segregation) and simplify our exposition, from now on we assume that $|N_i| = m$ for all $i = 1, ..., r$ (we discuss the implications of relaxing this assumption at the end of this section).

The complete characterization of the necessary and sufficient conditions for the fully segregated partition to be stable in the information sharing application directly exploits the characterization of the stable groups via the sequence $\{T_k^n\}_{k=0}^n$ developed in Propositions 1 and 4. In more detail, consider an agent a with taste parameter $t_i \in \{t_1, ..., t_r\}$. In the fully segregated partition such an agent belongs to the group N_i. To show stability, we need to consider all possible deviations of agent a to any set $N_j \cup \{a\}$, $j \neq i$. Since $|N_i| = m$ for all $i = 1, ..., r$, checking that deviations to the groups closest in tastes, $N_{i-1} \cup \{a\}$ and $N_{i+1} \cup \{a\}$, are not profitable is sufficient.[32] Thus, requiring a deviation from N_i to $N_{i+1} \cup \{a\}$ not to be profitable allows us to identify an upper bound for t_{i+1} for segregation to be stable. Similarly, requiring a deviation from N_i to $N_{i-1} \cup \{a\}$ not to be profitable allows us to identify a lower bound for t_{i-1} for segregation to be stable. Such bounds can be used to define an interval $\mathcal{T}_i$ around each t_i such that full segregation is stable if and only if *for any $i \neq j$, $t_j \notin \mathcal{T}_i$.* That is, t_i and t_j are sufficiently far from each other.

The idea that agents in other groups have sufficiently different tastes is then captured by them agreeing (among their cohesive groups of m agents) on very different allocations of signals. In particular, it can be shown that the intervals $\mathcal{T}_i$ correspond to *unions* of the original intervals in the sequence $\{T_k^m\}_{k=0}^m$. Figure 2 illustrates graphically the structure of the intervals $\mathcal{T}_i$ and $\mathcal{T}_{i+1}$ as unions of sets of the original sequence $\{T_k^m\}_{k=0}^m$.

We now turn to the robustness of the basic comparative statics obtained in Section 5.1. Recall that the main insight from Proposition 4 was that similarity in stable groups is stronger for extreme than for moderate tastes.[33] In the same spirit, in the finite population case we show that *segregation is easier to achieve for extreme types than for moderate ones.* Intuitively, suppose that agent a has taste parameter t_i, and consider a deviation from the group N_i to $N_{i+1} \cup \{a\}$. From Proposition 4, we know that the intervals $\{T_k^m\}_{k=0}^m$, each containing all the t's that correspond to the same optimal allocation of m signals, are narrower for extreme taste parameters and wider for moderate taste parameters. Thus, for a given distance between t_i and t_{i+1}, the disagreement between agent a and the agents in N_{i+1} is stronger if t_i and t_{i+1} are taste parameters closer to the extremes than if they are moderate.

---

[32] We have to consider just one potential deviation for agents with extreme tastes t_1 and t_r.

[33] In order to ease the description of our results, we assume that the sufficient conditions for the full sequence $\{T_k^m\}_{k=0}^m$ to follow the pattern described in Proposition 4 are met. These conditions are formally derived in the Appendix. They amount to q_A, q_B, and m being high enough.
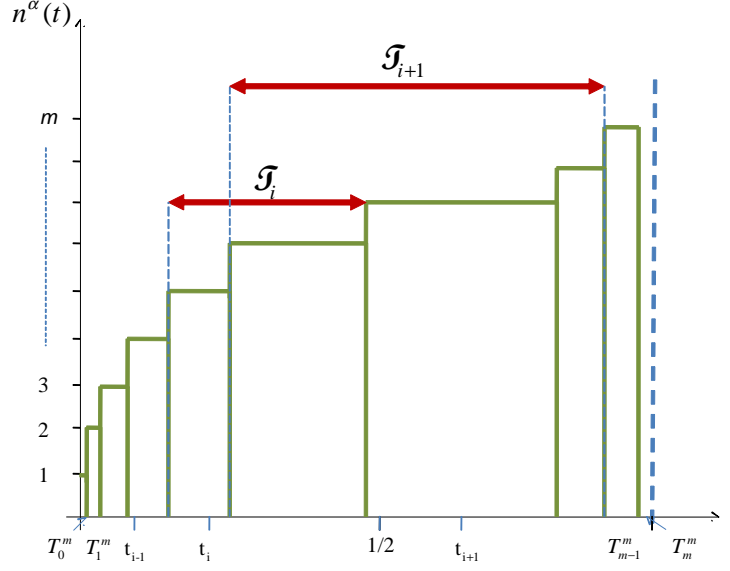
Figure 2: Segregation Intervals

In other words, the intervals $\{\mathcal{T}_i\}$ follow a similar pattern described in Section 5.1 for the interval sequence $\{T_k^m\}_{k=0}^m$.

This implies that, in information sharing environments, moderate individuals can be quite heterogeneous without segregation emerging, whereas extremists can be less heterogeneous in comparison and still allow for multiple small and homogeneous groups. In particular, in the case of a finite population uniformly distributed on equidistant points on the interval $[0, 1]$, the most segregated stable partition that can emerge will tend to display *large*, heterogeneous groups of moderate agents and multiple, *smaller* groups of extremists. Proposition 5 formalizes the above discussion.

Proposition 5 (Information Sharing - Stable Segregation) *For any $t_i \in T$ there exists an interval $\mathcal{T}_i$ such that full segregation is a stable partition if and only if for any $t_i, t_j \in T$, t_i and t_j are sufficiently far from each other: $t_j \notin \mathcal{T}_i$. Furthermore, full segregation is easier to sustain as stable for extreme tastes than for moderate tastes: there exists i such that the intervals $\mathcal{T}_1, \mathcal{T}_2, ..., \mathcal{T}_i$ are increasing in length; similarly, there exists j such that the intervals $\mathcal{T}_j, \mathcal{T}_{j+1}, ..., \mathcal{T}_s$ are decreasing in length.*

To conclude this section, we address the implications of relaxing the assumption that $|N_i| = m$ for all i, and allowing for the sets of agents sharing the same taste parameter to be of different sizes. Note that agents of the same taste will always want to be in the same group, following the lines of Proposition 1 above. Now, the larger the number of individuals of a particular taste parameter is, the more appealing the group that contains these individuals becomes (by the sheer volume of information they collect). Therefore, in order to achieve stable segregation, the other agents need to be further away in terms of their preferences. So, the band around a particular type that guarantees segregation increases in width with the number of agents of that type. This result can be applied to derive straightforward comparative statics with respect to the population distribution. For example, consider two cumulative distributions over tastes in the population G and G' such that G is a mean-preserving spread of G'. In this case, the most segregated partition corresponding to G' will be characterized by *larger and more heterogeneous* groups of moderates, and *smaller and more fractioned* groups of extremists relative to the most segregated partition corresponding to G.

6. Conclusions

The model developed in this paper addresses the properties of peer groups that arise in equilibrium when individuals have different tastes and make contributions to public projects. There are three main results that emerge from our analysis. First, sorting arises in equilibrium – stable groups are composed of individuals that are sufficiently similar in tastes. Second, stable groups may remain heterogeneous even when the population grows – in particular, they do not necessarily become so refined that they contain only one type of individual. Finally, when the population is small, it is natural to consider stable partitions of the population into groups. We show that stable partitions are 'consecutive', whenever two types belong to the same partition element, so do all types that are ordered between them.

The model can naturally be applied to an information-sharing environment, in which members of a group share privately collected information. In this setting, the similarity observed among members of stable groups is more pronounced for extreme agents than for moderate ones. Furthermore, in small populations, full segregation is easier to achieve for agents with extreme tastes than for agents with moderate ones.

Our basic model can be investigated further in several directions. Throughout the analysis we assume that contributions are free, and, for the infinite population case, we assume that group size is exogenously fixed (this last assumption allows us to isolate one aspect of choice, that of groups' composition). There are then two natural directions in which the analysis can be extended. The first has to do with the implications of introducing contribution costs. When contribution costs are low, the characterization of stable groups should coincide with the characterization we provide here. However, as contribution costs become large, agents of extremely different tastes could coexist in stable groups since this heterogeneity provides stronger incentives to contribute and mitigates the free-rider problem. This would suggest that, as technologies improve, stable groups exhibit more similarity in tastes. The second direction has to do with the correlation between group composition and group size. For instance, if each individual member incurred a connection cost (that depended on the number of her connections), one could extend our stability concept to make the size of groups a second object of choice. Namely, individuals would choose optimally both the type and the number of their connections.

7.   Appendix

Proof of Lemma 1. Since $\frac{\partial^2 U}{\partial t \partial k_A} > 0$ and $\frac{\partial^2 U}{\partial t \partial k_B} < 0$, if an agent of taste t prefers making an A-contribution over a B-contribution, so would any agent of taste $t' > t$. Similarly, if an agent of taste t prefers making a B-contribution, so would any agent of taste $t' < t$. Let k^* be defined as the maximal $k \in \{0, 1, ..., n\}$ for which agent k *weakly* prefers making an A-contribution over a B-contribution. That is, the maximal $k \in \{0, 1, ..., n\}$ for which

$$U(t_k, k, n - k) \geq U(t_k, k - 1, n - k + 1). \tag{5}$$

If (5) is not satisfied for any agent in the group (i.e., $U(t_1, 0, n) > U(t_1, 1, n - 1)$), then $k^* = 0$. Given our tie-breaking rule, k^* clearly defines a monotonic equilibrium as prescribed.

In order to show uniqueness of the amount of A-contributions made in equilibrium, we show that if $(x_1, ..., x_n)$ is a pure equilibrium such that for some $i > j$, $x_i = A$ and $x_j = B$, then $(y_1, ..., y_n) \in \{A, B\}^n$, where $y_l = x_l$ for all $l \neq i, j$, $y_i = A$, and $y_j = B$ constitutes a pure equilibrium as well.

Indeed, assume that

$$n^A = |\{x_l = A, l \neq i, j\}| \quad \text{and} \quad n^B = |\{x_l = B, l \neq i, j\}|.$$

$(x_1, ..., x_n)$ consisting an equilibrium requires that agent i best responds. In particular, $U(t_i, n^A + 1, n^B + 1) \geq U(t_i, n^A + 2, n^B)$. Similarly, agent j best responding requires that $U(t_j, n^A + 1, n^B + 1) \geq U(t_j, n^A, n^B + 2)$. Since $\frac{\partial^2 U(t, k, n-k)}{\partial t \partial k} > 0$, the above best response restrictions hold for agents i and j under the profile $(y_1, ..., y_n)$ as well, while all other players' best responses remain unchanged. The claim follows. ∎

Proof of Proposition 1.

Step 1. We first show that, given a group size n, for any taste $t \in [0, 1]$, there exist $l(t), h(t) \in [0, 1]$, $l(t) \leq t \leq h(t)$, such that any group of agents with tastes $t_1 \geq t_2 \geq ... \geq t_n$ (one of whom is t) is optimal if and only if

$$t_1 \geq t_2 ... \geq t_{n^A(t)} \geq l(t) \quad \text{and} \quad h(t) > t_{n^A(t)+1} \geq ... \geq t_n.$$

Indeed, for any taste t, an optimal choice for a group entails choosing $n^A(t)$ agents who make an A-contribution and $n^B = n - n^A(t)$ agents who make a B-contribution. From Lemma 1, this is tantamount to choosing a group of agents with tastes $t_1 \geq t_2 \geq ... \geq t_n$ (of which agent t is a member) such that condition (5) holds only for $i = 1, .., n^A(t)$.

Define

$$\Delta \equiv U(t, k, n - k) - U(t, k - 1, n - k + 1).$$

Since $\frac{\partial^2 U(t, k, n-k)}{\partial t \partial k} > 0$, it follows that:

$$\frac{\partial \Delta}{\partial t} = \frac{\partial U(t, k, n - k)}{\partial t} - \frac{\partial U(t, k - 1, n - k + 1)}{\partial t} > 0.$$

Therefore, the conditions for the group of agents with tastes $t_1 \geq t_2 \geq ... \geq t_n$ to be optimal can be described as follows.

1. For $i = 1, .., n^A(t)$,

$$U(t_i, n^A(t), n^B(t)) \geq U(t_i, n^A(t) - 1, n^B(t) + 1).$$

If this inequality holds for any $t_i \leq t$, define $l(t) = 0$. Otherwise, define $l(t)$ so that

$$U(l(t), n^A(t), n^B(t)) = U(l(t), n^A(t) - 1, n^B(t) + 1)$$

and $l(t)$ is the preference parameter of an agent who is precisely indifferent between allocations prescribing $n^A(t)$ or $n^A(t) - 1$ contributions directed at task A. The condition then translates to $t_1 \geq ... \geq t_{n^A(t)} \geq l(t)$.

2. For $i = n^A(t) + 1, .., n$,

$$U(t_i, n^A(t), n^B(t)) \geq U(t_i, n^A(t) + 1, n^B(t) - 1).$$

In analogy to the above, if this inequality holds for all $t_i \geq t$, define $h(t) = 1$. Otherwise, define $h(t)$ so that

$$U(h(t), n^A(t), n^B(t)) = U(h(t), n^A(t) + 1, n^B(t) - 1)$$

and $h(t)$ is the preference parameter of an agent who is precisely indifferent between allocations prescribing $n^A(t)$ or $n^A(t) + 1$ contributions directed at task A. The condition then translates to $h(t) > t_{n^A(t)+1} \geq ... \geq t_n$.

Step 2. We first show that $m^A(t)$ is increasing in t. Indeed, suppose $t' > t$. Since $\frac{\partial^2 U(t,k,n-k)}{\partial t \partial k} > 0$ it follows that:

$$\frac{\partial U(t', k, n - k)}{\partial k}\bigg|_{k=m^A(t)} > \frac{\partial U(t, k, n - k)}{\partial k}\bigg|_{k=m^A(t)} = 0.$$

Since $\frac{\partial U(t',k,n-k)}{\partial k}\bigg|_{k=m^A(t')} = 0$ and

$$\frac{\partial^2 U(t, k, n - k)}{\partial k^2} = \frac{\partial^2 U(t, k, n - k)}{\partial k_A^2} + \frac{\partial^2 U(t, k, n - k)}{\partial k_B^2} < 0,$$

we get that $m^A(t') > m^A(t)$.

If $n^A(0) = n^A(1)$ then all agents agree on the allocation of contributions and any group of n agents

is stable. That is, the partition containing one interval $[0, 1]$ identifies stable groups.

Suppose $n^A(0) < n^A(1)$. For any $k = n^A(0) + 1, ..., n^A(1)$, denote by $t(k)$ the taste parameter with which an agent is indifferent between $k - 1$ and k contributions to task A. That is,

$$U(t(k), k - 1, n - k + 1) = U(t(k), k, n - k).$$

Monotonicity of $m^A(t)$ insures that $t(k)$ is defined uniquely for all k.

It follows that an agent with taste parameter t would like $n^A(0)$ contributions to be directed to task A whenever $t \in [0, t(n^A(0) + 1))$, would like $n^A(1)$ units to be directed at task A whenever $t \in [t(n^A(1)), 1]$, and would like $n^A(t) = k \in \{n^A(0) + 1, ..., n^A(1) - 1\}$ contributions to be directed at task A if and only if $t \in [t(k), t(k + 1))$. Since, from Step 1, each agent can achieve her optimal allocation of contributions with some group, stability boils down to all of the group members agreeing on the ideal amount of contributions directed at each task. The claim follows. ∎

Proof of Proposition 2. Recall that the intervals $\{T_k^n\}_v^{v'}$ are defined as $T_k^n \equiv \{t \mid n^A(t) = k\}$. Since $n^A(t) = \lfloor m^A(t) \rfloor$ if $m^A(t) \in [0, n]$, the intervals' extremes are identified with the projections of integer numbers onto the intervals $[0, 1]$ via the function $m^A(t)$. Let $\frac{dm^A(t)}{dt} \leq M$ for all $t \in [a, b]$. By construction, types in $[a, b]$ can belong to at most $L \equiv \lceil M(b - a) \rceil + 1$ stable groups for any group size n. ∎

Proof of Proposition 3.

1. Recall that $\{t_1, .., t_r\}$ is the set of taste parameters t_i such that there is at least one agent in N with taste t_i. Let any group of agents $G \subseteq N$ be identified by a vector $(z_1, .., z_r)$, where $z_l \leq m_l$ is the number of agents of taste t_l in group G. Suppose that $\mathcal{G} = \{G_1, .., G_s\}$ is a stable allocation, and let $t_i, t_j, t_h \in \{t_1, .., t_r\}$ be such that $t_i > t_j > t_h$, and suppose that two agents with tastes t_i and t_h, respectively, are in a group $G \in \mathcal{G}$ and that an agent with taste t_j is in $G' \neq G$. Assume that in any equilibrium of the task-selection game, k_A contributions are directed at task A and k_B contributions are directed at task B in group G, while k'_A contributions are directed at task A and k'_B contributions are directed at task B in group G'. Since an agent with taste t_j at least weakly prefers G' over G it must be the case that either $k_A < k'_A$ or $k_B < k'_B$. Suppose $k_A < k'_A$. Certainly, if $k_B \leq k'_B$, the other

two agents (of types t_i and t_h) would benefit by switching to group G'. Assume then that $k_B > k'_B$. Optimality for the agent of type t_j then implies:

$$U(t_j, k'_A, k'_B) - U(t_j, k_A, k_B) \geq 0.$$

Since $\frac{\partial^2 U}{\partial t \partial k_A} > 0$,

$$U(t_j, k'_A, k'_B) - U(t_j, k_A, k'_B) \leq U(t_i, k'_A, k'_B) - U(t_i, k_A, k'_B)$$

and, similarly, since $\frac{\partial^2 U}{\partial t \partial k_B} < 0$,

$$U(t_j, k_A, k_B) - U(t_j, k_A, k'_B) \geq U(t_i, k_A, k_B) - U(t_i, k_A, k'_B).$$

It follows that:

$$
\begin{aligned}
&U(t_i, k'_A, k'_B) - U(t_i, k_A, k_B) = \\
&[U(t_i, k'_A, k'_B) - U(t_i, k_A, k'_B)] - [U(t_i, k_A, k_B) - U(t_i, k_A, k'_B)] \geq \\
&[U(t_j, k'_A, k'_B) - U(t_j, k_A, k'_B)] - [U(t_j, k_A, k_B) - U(t_j, k_A, k'_B)] = \\
&U(t_j, k'_A, k'_B) - U(t_j, k_A, k_B) \geq 0.
\end{aligned}
$$

Since the agent of type t_i joining G' would imply an additional available contribution, it follows that an agent of type t_i would strictly benefit by shifting from group G to group G', in contradiction. The case $k_A > k'_A$ follows analogously.

2. Suppose that $\mathcal{G}$ is a stable allocation, and suppose that $G, G' \in \mathcal{G}$, $G \neq G'$ both contain at least one agent of taste t_i. If one group identified by $(x_1, .., x_r)$ is (weakly) preferred by an agent of taste t_i to the group identified by $(x'_1, .., x'_r)$, we write $(x_1, .., x_r) \succcurlyeq_{t_i} (x'_1, .., x'_r)$. Assume that G is identified by $(x_1, .., x_i, ..x_r)$ and that G' is identified by $(y_1, .., y_i, .., y_r)$. For an agent of taste t_i, a deviation from G to G' is unprofitable if $(x_1, .., x_i, ..x_r) \succcurlyeq_{t_i} (y_1, .., y_i + 1, .., y_r)$. Similarly, for an agent of taste t'_i, a deviation from G' to G is unprofitable if $(y_1, .., y_i, .., y_r) \succcurlyeq_{t_i} (x_1, .., x_i + 1, .., x_r)$. However, any agent of taste t_i strictly benefits from having her group augmented by one more member of her own type. Since $(z_1, .., z_s + 1, .., z_r) \succ_{t_s} (z_1, .., z_s, .., z_r)$, for any $(z_1, .., z_s, .., z_r)$ and t_s, we get a contradiction. ∎

Proof of Proposition 4. From the definition of $n^A(t)$, in the information-collection setting, we have $n^A(t) = \lfloor m^A(t) \rfloor$ if $m^A(t) \in [0, n]$, $n^A(t) = 0$ if $m^A(t) < 0$, and $n^A(t) = n$ if $m^A(t) > n$, where

$m^A(t)$ is the number achieving equality in condition (1). Simple algebraic manipulation yields:

$$m^A(t) = \frac{\ln\left(\frac{1}{t} - 1\right) + n\ln(1 - q_B) + \ln(1 - q_A) + \ln\frac{q_B}{q_A}}{\ln(1 - q_A) + \ln(1 - q_B)}.$$

Differentiating $m^A(t)$ we get:

$$\frac{dm^A(t)}{dt} = -\frac{1}{t(1-t)\ln[(1-q_A)(1-q_B)]} > 0;$$
$$\frac{d^2 m^A(t)}{dt^2} = \frac{1-2t}{[t(1-t)]^2 \ln[(1-q_A)(1-q_B)]}.$$

Therefore, $m^A(t)$ is an increasing function that is concave up to $t = \frac{1}{2}$ and convex thereafter. Since for any $k = 1, ..., n-1$, $T_k^n = \left(m^A\right)^{-1}([k, k+1))$, this implies that the sequence of intervals $\{T_k^n\}_{k=1}^{n-1}$ is such that the intervals are increasing in length until the interval $T_{\hat{k}}^n$ such that $1/2 \in T_{\hat{k}}^n$ and decreasing thereafter.

We now address the extreme intervals T_0^n and T_n^n. We will show that these intervals follow the same pattern of $\{T_k^n\}_{k=1}^{n-1}$ if either q_A, q_B are high enough, or n is high enough. Recall that $T_0^n = [0, t(1))$ and $T_n^n = [t(n), 1]$. From the definition of $t(k)$ above, we get

$$t(k) = \frac{(1 - q_B)^{n-k} q_B}{(1 - q_A)^{k-1} q_A + (1 - q_B)^{n-k} q_B}. \tag{6}$$

We have that $t(1) \leq t(2) - t(1)$ (i.e., interval T_0^n is shorter than T_1^n) if and only if

$$\frac{2(1 - q_B)^{n-1} q_B}{q_A + (1 - q_B)^{n-1} q_B} \leq \frac{(1 - q_B)^{n-2} q_B}{(1 - q_A) q_A + (1 - q_B)^{n-2} q_B}.$$

Rearranging, the above condition is satisfied if and only if

$$2q_A(1 - q_A)(1 - q_B) + (1 - q_B)^{n-1} q_B \leq q_A. \tag{7}$$

Condition (7) is a necessary and sufficient condition on q_A, q_B, and n such that the interval T_0^n follows the same pattern of the sequence $\{T_k^n\}_{k=1}^{n-1}$. Note that, since $(1 - x)x$ is maximized in $\left[\frac{1}{2}, 1\right]$ at $x = 1/2$, $q_A, q_B \geq \frac{1}{2}$ is a sufficient condition to guarantee (7), as we have

$$2(1 - q_A) q_A(1 - q_B) + (1 - q_B)^{n-1} q_B \leq (1 - q_A) q_A + \frac{1}{4} \leq \frac{1}{2} \leq q_A.$$

Moreover, if $(1 - q_A)(1 - q_B) < 1/2$, condition (7) holds for n large enough. The interval T_n^n is shorter than T_{n-1}^n if and only if $1 - t(n) \leq t(n) - t(n-1)$. After rearranging, this is equivalent to

$$2q_B(1 - q_A)(1 - q_B) + (1 - q_A)^{n-1} q_A \leq q_B. \tag{8}$$

As before, it is easy to see that if $(1 - q_A)(1 - q_B) < 1/2$, condition (8) is satisfied for large enough n, and that $q_A, q_B \geq \frac{1}{2}$ is a sufficient condition for (8) to be satisfied for any n. $\blacksquare$

Proof of Corollary 1. The first part follows directly from Proposition 2 since in this setting $\frac{dm^A}{dt}$ is uniformly bounded for any interval $[a, b] \subset (0, 1)$. For the second part, observe that the interval T_0^n contains all t such that $t < \frac{(1-q_B)^{n-1}q_B}{q_A + (1-q_B)^{n-1}q_B} = t^n(1)$. It follows that $T_0^{n'} \subsetneq T_0^n$ for any $n' > n$ and, for any $q_A, q_B \in (0, 1)$, $t^n(1) \searrow_{n \to \infty} 0$. so the interval T_0^n shrinks to a singleton as the size of the group becomes infinitely large. Similarly, T_n^n contains all $t \geq \frac{q_B}{(1-q_A)^{n-1}q_A + q_B} = t^n(n)$. It follows that $T_{n'}^{n'} \subsetneq T_n^n$ for any $n' > n$ and, for any $q_A, q_B \in (0, 1)$, $t^n(n) \nearrow_{n \to \infty} 1$, so the interval T_n^n shrinks to a singleton as the size of the group becomes infinitely large as well. $\blacksquare$

Proof of Proposition 5. For $x = A, B$, denote by $z^x(t, h) \equiv n^x(t)$ the optimal number of x-signals out of a total of h signals for an agent of taste t. Moreover, for any $t_1, t_2 \in [0, 1]$, let $w^x(t_1, t_2, h_1, h_2)$ denote the equilibrium number of x-signals collected in a group that is composed of h_1 agents of taste t_1 and h_2 agents of taste t_2 (well-defined from Lemma 1).

Observe that if $z^A(t_1, N) > z^A(t_2, N)$ then either $w^A(t_1, t_2, 1, N) = z^A(t_2, N) + 1$ (if $z^A(t_2, N + 1) = z^A(t_2, N) + 1$), or $w^A(t_1, t_2, 1, N) = z^A(t_2, N)$ (if $z^A(t_2, N + 1) = z^A(t_2, N)$).

Consider a fully segregated partition and suppose that agent $a \in N_i$ has taste parameter t_i for $i \in \{1, .., r\}$. Since $|N_i| = m$ for all i, checking that such an agent does not have a profitable deviation by joining $N_{i+1} \cup \{a\}$ and $N_{i-1} \cup \{a\}$ is enough to guarantee that this agent does not have profitable deviations (note that for $i = 1, r$, there is only one constraint to check). Consider a deviation of agent a from N_i to $N_{i+1} \cup \{a\}$. Since $t_i \geq t_{i+1}$, $z^A(t_i, n) \geq z^A(t_{i+1}, n)$ for all $n \geq 1$. A necessary condition for the deviation not to be strictly beneficial is that $z^A(t_i, m) > z^A(t_{i+1}, m)$. Suppose first

that $w^A(t_i, t_{i+1}, 1, m) = z^A(t_{i+1}, m) + 1$. It follows that the deviation is not profitable whenever

$$t_i \left[1 - \frac{1}{2}(1-q_A)^{z^A(t_i,m)}\right] + (1-t_i)\left[1 - \frac{1}{2}(1-q_B)^{m-z^A(t_i,m)}\right] \tag{9}$$
$$\geq t_i \left[1 - \frac{1}{2}(1-q_A)^{z^A(t_{i+1},m)+1}\right] + (1-t_i)\left[1 - \frac{1}{2}(1-q_B)^{m-z^A(t_{i+1},m)}\right],$$

or, rearranging terms,

$$t_i \left[(1-q_A)^{z^A(t_{i+1},m)+1} - (1-q_A)^{z^A(t_i,m)}\right] \geq (1-t_i)\left[(1-q_B)^{m-z^A(t_i,m)} - (1-q_B)^{m-z^A(t_{i+1},m)}\right]. \tag{10}$$

For $x = A, B$, define $v^x(k) \equiv (1-q_x)^k - (1-q_x)^{k+1}$, so that $v^A(k)$ is the marginal contribution of the $(k+1)$'th A-signal and similarly for $v^B(k)$ (up to a factor of $\frac{1}{2}$). Note that, for $x = A, B$, $v^x(k)$ is decreasing in k. Thus, substituting and rearranging terms, we can rewrite (10) as follows:

$$\frac{t_i}{1-t_i} \geq \frac{\displaystyle\sum_{k=m-z^A(t_i,m)}^{m-z^A(t_{i+1},m)-1} v^B(k)}{\displaystyle\sum_{k=z^A(t_{i+1},m)+1}^{z^A(t_i,m)-1} v^A(k)}, \tag{11}$$

where we use the convention that $\sum_{k=w}^{w-1} v^A(k) \equiv 0$ for any w. Condition (11) implicitly defines a condition on t_i and t_{i+1} for a deviation from N_i to N_{i+1} to be unprofitable. Observe that, if $w^A(t_i, t_{i+1}, 1, m) = z^A(t_{i+1}, m)$, a deviation of agent a of taste parameter t_i to a group of m agents of taste parameter t_{i+1} is less profitable than a deviation to a group in which $z^A(t_{i+1}, m) + 1$ out of $m+1$ agents collect the A-signal, and therefore, condition (11) is sufficient to guarantee that such deviation is not profitable.

If t_i is fixed, condition (11) is weaker the lower is t_{i+1} (that is, the further apart t_i and t_{i+1} are). This guarantees that there exists $\underline{t}(t_i)$ such that a deviation of agent a in N_i to N_{i+1} is unprofitable if and only if $t_{i+1} < \underline{t}(t_i)$. If $t_{i+1} \in T_0^m$ and condition (11) is not satisfied, then $\underline{t}(t_i) = 0$. Set $\underline{t}(t_i) = 0$ if $z^A(t_i, m) = z^A(t_{i+1}, m) = 0$.

We can follow a similar procedure by considering a deviation from N_i to $N_{i-1} \cup \{a\}$ and defining a taste $\bar{t}(t_i)$ such that a deviation from N_i to $N_{i-1} \cup \{a\}$ is unprofitable if and only if $t_{i-1} > \bar{t}(t_i)$. The intervals $\{\mathcal{T}_i\}_{i=1}^r$ are obtained by setting for any $i \in \{1, .., r\}$, $\mathcal{T}_i \equiv [\underline{t}(t_i), \bar{t}(t_i))$ whenever $\bar{t}(t_i) <$

1 and $\mathcal{T}_i \equiv [\underline{t}(t_i), 1]$ whenever $\bar{t}(t_i) = 1$. To see that $\mathcal{T}_i$ are unions of contiguous intervals of the sequence $\{T_k^m\}_{k=1}^m$, observe that all t_{i+1} in the same interval T_k^m share the same optimal allocation $(z^A(t_{i+1}, m), z^B(t_{i+1}, m))$. Thus, for a given t_i, if condition (11) is satisfied for $t_{i+1} \in T_k^m$, then it is satisfied for any $t'_{i+1} \in T_k^m$.

We now show the comparative statics of the intervals $\{\mathcal{T}_i\}_i$. For each i, denote $\mathcal{T}_i^- = \mathcal{T}_i \cap [0, t_i]$ and $\mathcal{T}_i^+ = \mathcal{T}_i \cap [t_i, 1]$ the sub-intervals of $\mathcal{T}_i$ that are to the left and right of t_i, respectively. Notice that for sufficiently low t_i, $\mathcal{T}_i^- = [0, t_i]$, the length of which is increasing in t_i (and decreasing in i). Similarly, for sufficiently high t_i, $\mathcal{T}_i^+ = [t_i, 1]$, the length of which is decreasing in t_i (and increasing in i). To show the claim, it suffices to illustrate that for sufficiently low t_i, $\mathcal{T}_i^+, ..., \mathcal{T}_s^+$ are decreasing in length and, similarly, for sufficiently high t_i, $\mathcal{T}_1^-, ..., \mathcal{T}_i^-$ are increasing in length. Consider t and t', with $t' > t$, and $k \in \{1, .., z^A(t, m)\}$, such that the agent of type t prefers to stay in her group collecting $z^A(t, m)$ A-signals than to be the $m + 1$'th member of a group in which, without her, $z^A(t, m) - k$ A-signals are collected. Upon joining such a group, the agent would be collecting an A-signal, and so the corresponding incentive constraint (similar to condition (10)) implies that:

$$\frac{t}{1-t} \geq \frac{(1-q_B)^{m-z^A(t,m)} - (1-q_B)^{m-z^A(t,m)+k}}{(1-q_A)^{z^A(t,m)-k+1} - (1-q_A)^{z^A(t,m)}}. \tag{12}$$

Let

$$\mathcal{A}(t) \equiv t \left[(1-q_A)^{m^A(t)-k+1} - (1-q_A)^{m^A(t)} \right],$$
$$\mathcal{B}(t) \equiv (1-t) \left[(1-q_B)^{m-m^A(t)} - (1-q_B)^{m-m^A(t)+k} \right]$$

and note that whenever $z^A(t, m) = z^A(t', m)$, if (12) holds for t, it will hold for t'. In order to show the claim, it is therefore sufficient to focus on t and t' that are at the cusps of our original intervals $\{T_k^m\}_k$, for which $m^A(t) = z^A(t, m)$ and $m^A(t') = z^A(t', m)$. For such t and t', (12) is satisfied whenever $\mathcal{A}(t) \geq \mathcal{B}(t)$ and $\mathcal{A}(t') \geq \mathcal{B}(t')$, respectively. From the derivations in the proof of Proposition 5, recall that

$$\frac{dm^A(t)}{dt} = -\frac{1}{t(1-t)\ln\left[(1-q_A)(1-q_B)\right]} > 0.$$

Simple algebraic manipulations imply that $\mathcal{A}'(t)$ is positive if $1 + t \ln(1 - q_A) \frac{dm^A(t)}{dt} > 0$, which, after further manipulation, is equivalent to

$$t < \frac{\ln(1 - q_B)}{\ln(1 - q_A) + \ln(1 - q_B)}.$$

Similarly, it is easy to show that $\mathcal{B}'(t)$ positive if $1 + (1 - t) \ln(1 - q_B) \frac{dm^A(t)}{dt} < 0$, or

$$t < \frac{\ln(1 - q_B)}{\ln(1 - q_A) + \ln(1 - q_B)}.$$

Note that $\mathcal{A}(0) = 0$ and $\mathcal{A}(1) = \left[(1 - q_A)^{m-k+1} - (1 - q_A)^m\right] > 0$, while $\mathcal{B}(0) = (1 - q_B)^m - (1 - q_B)^{m+k} > 0$ and $\mathcal{B}(1) = 0$. Since both $\mathcal{A}(t)$ and $\mathcal{B}(t)$ are decreasing for any $t > \frac{\ln(1-q_B)}{\ln(1-q_A)+\ln(1-q_B)}$, and $\mathcal{A}(1) > \mathcal{B}(1)$, it must be the case that for t high enough, $\mathcal{A}(t) \geq \mathcal{B}(t)$ implies $\mathcal{A}(t') \geq \mathcal{B}(t')$ whenever $t' > t$ (in addition, it is easy to show that both $\mathcal{A}(t)$ and $\mathcal{B}(t)$ are concave; therefore, they can cross at most once for $t > \frac{\ln(1-q_B)}{\ln(1-q_A)+\ln(1-q_B)}$). From the comparative statics of Proposition 4, it follows that for any t_i high enough, $\mathcal{T}_1^-, ..., \mathcal{T}_i^-$ are increasing in length, as was required. The case of low t_i follows analogously. ∎

8. REFERENCES

Benhabib, Jess, Alberto Bisin, and Matthew O. Jackson (editors) (2010), *The Handbook of Social Economics*, Elsevier.

Bramoullé, Yann, Sergio Currarini, Matthew O. Jackson, Paolo Pin, and Brian W. Rogers (2012), "Homophily and Long-Run Integration in Social Networks," *Journal of Economic Theory*, 147(5), 1754-1786.

Bogomolnaia, Anna, Michel Le Breton, Alexei Savvateev, and Shlomo Weber (2007), "Stability under Unanimous Consent, Free Mobility and Core," *International Journal of Game Theory*, 35(2), 185-204.

Calvó-Armengol Antoni, Joan De Martí, and Andrea Prat (2011), "Communication and Influence," mimeo.

Coleman, James S. (1966), *Medical Innovation: A Diffusion Study*, Second Edition, Bobbs-Merrill. New York.

Conley, Timothy G. and Christopher R.. Udry (2010), "Learning About a New Technology: Pineapple in Ghana," *The American Economic Review*, 100(1), 35-69.

Currarini, Sergio, Matthew O. Jackson, and Paolo Pin (2009), "An Economic Model of Friendship: Homophily, Minorities and Segregation," *Econometrica*, 77(4), 1003–1045.

Demange, Gabrielle (1994), "Intermediate Preferences and Stable Coalition Structures," *Journal of Mathematical Economics*, 23, 45-58.

Dunbar, Robin I. M. (1992), "Neocortex Size as a Constraint on Group Size in Primates," *Journal of Human Evolution*, 22, 469-493.

Ellickson, Bryan, Birgit Grodal, Suzanne Scotchmer, and William R. Zame (1999), "Clubs and the Market," *Econometrica*, 67, 1185-1218.

Evans, William N., Wallace E. Oates, and Robert M. Schwab (1992), "Measuring Peer Group Effects: A Study of Teenage Behavior," *The Journal of Political Economy*, 100(5), 966-991.

Foster, Andrew D. and Mark R. Rosenzweig (1995), "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *The Journal of Political Economy*, 103(6), 1176-1209.

Glaeser, Edward L., Bruce Sacerdote, and Jose A. Scheinkman (1996), "Crime and Social Interaction," *The Quarterly Journal of Economics*, 111(2), 507-548.

Goeree, Jacob, Margaret A. McConnell, Tiffany Mitchell, Tracey Tromp, and Leeat Yariv (2010), "The 1/d Law of Giving," *American Economic Journal: Microeconomics*, 2(1), 183-203.

Granovetter, Mark (1994), *Getting a Job: A Study of Contacts and Careers*, Northwestern University Press. Evanston.

Greenberg, Joseph, and Shlomo Weber (1986), "Strong Tiebout Equilibrium under Restricted Preferences Domain," *Journal of Economic Theory*, 38, 101-117.

Huckfeldt, Robert R. and John Sprague (1995), *Citizens, Politics and Social Communication*, Cambridge Studies in Public Opinion and Political Psychology.

Kandell, Denise B. (1978), "Homophily, Selection and Socialization in Adolescent Friendships," *The American Journal of Sociology*, 84(2), 427-436.

Katz, Elihu and Paul F. Lazarsfeld (1955), *Personal influence: The Part Played by People in the Flow of Mass Communication*, Glencoe, IL: Free Press.

Lazarsfeld, Paul F.and Robert K. Merton (1954), "Friendship as a social process: a substantive and methodological analysis," in *Freedom and Control in Modern Society*, ed. M Berger, 18–66. New York: Van Nostrand.

Lipsman, Andrew (2008), "Social Networking Explodes Worldwide as Sites Increase Their Focus on Cultural Relevance," *comScore*, 12/3/2008.

Lynd, Robert S., and Helen M. Lynd (1929), *Middletown: A Study in Contemporary American Culture*, New York, Harcourt, Brace and Company, New York, NY.

Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60(3), 531-542.

Manski, Charles F. (2000), "Economic Analysis of Social Interactions," *The Journal of Economic Perspectives*, 14(3), 115-136.

Marmaros, David and Bruce Sacerdote (2006), "How do Friendships Form?," *The Quarterly Journal of Economics*, 121(1), 79-119.

Mayer, Adalbert and Steven L. Puller (2008), "The Old Boy (and Girl) Network: Social Network Formation on University Campuses," *Journal of Public Economics*, 92(1-2), 329-347.

McCarty, Christopher, Peter D. Killworth, H. Russell Bernard, Eugene C. Johnsen, and Gene A. Shelley (2000), "Comparing Two Methods for Estimating Network Size," *Human Organization*, 60, 28-39.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook (2001), "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415-444.

Myers, David G. (2007), *Social Psychology*, McGraw-Hill, New York, NY.

Peski, Marcin (2008), "Complementarities, Group Formation and Preferences for Similarity," mimeo.

Thomas, Timothy L. (2009), "Countering Internet Terrorism," *IO Sphere*.

Tiebout, Charles M. (1956), "A Theory of Local Expenditures," *The Journal of Political Economy*, 64(5), 416-424.

Topa, Giorgio (2001), "Social Interactions, Local Spillovers and Unemployment," *The Review of Economic Studies*, 68(2), 261-295.

Wooders, Myrna, Edward Cartwright, and Reinhard Selten (2006), "Behavioral Conformity in Games with Many Players," *Games and Economic Behavior*, 57, 347-360.