Appendix of “Homophily in Peer Groups”
The Costly Information Case

Mariagiovanna Baccara   Leeat Yariv

August 19, 2012

1 Introduction

In this Appendix we study the information sharing application analyzed in Section 5 of the paper when information is costly. This is relevant to many environments, ranging from focused wine or book clubs (in which participants have to jointly provide resources, be it money or time), to computer programming forums, whose members decide whether to invest in learning a new coding strategy.

When information gathering becomes costly, a free rider problem can arise within groups that are stable absent information costs. Since more extreme individuals have greater incentives to acquire information on the issue they care about, introducing group polarization has the benefit of mitigating the free rider problem by weakening the incentive constraints in the information-collection stage.

We show that when groups are small, the free rider problem is not severe, and stable groups coincide with those we describe for the free-information case in the paper. As the group size increases, the free rider problem becomes acute and, for sufficiently large groups, stability entails *extreme preference polarization*. Nonetheless, for intermediate group sizes, two types of groups are stable. First, we show that some of the (sufficiently homogeneous) stable groups in the free information costs case survive as stable. Second, we characterize stable *mildly polarized groups*. These groups include some extremists, who collect information on their preferred issue, and some moderates, who are close enough in their preferences to agree on the optimal allocation of the number of signals collected across issues.[1]

Since the driving force behind the appearance of stable polarization is the free rider problem, reducing

[1]The idea that agents’ preferences may alleviate incentive constraints in collective settings with costly information appears in some recent mechanism design literature, see Che and Kartik (2009) and Gerardi and Yariv (2008).

the cost of information collection plays a similar role to a reduction of group size: they both alleviate the free rider problem. A decrease in the cost of information may occur with the development of a new technology for information gathering (such as the Internet, or a search engine). Thus, our results suggest that, *as technology improves, stable groups exhibit more similarity in tastes*. Indeed, when information costs are high, polarized individuals are necessary for a group to gather information on both issues. As information costs decrease, information gathering becomes feasible for moderate individuals as well. Thus, agreement on the optimal way to go about collecting information becomes the prominent criterion around which stable groups form. This is consistent with a large body of empirical work studying the effects of new information technologies on social affiliations. For instance, Sproull and Kiesler (1991) depicted how the introduction of the telephone affected connection between similar individuals. Recently, Rosenblat and Mobius (2004) studied coauthored papers in top economics journals between the years 1969-1999. They showed how the introduction of the Internet in the early 1990's is linked with a 20% decrease in the realization of projects with a dissimilar coauthor.

2 The Model

We assume that observing a signal from either source comes at a cost of $c > 0$, which we assume to be entirely born by the agent who collects the signal. This corresponds to any context in which expertise requires effort (e.g., following the media, learning a new software, etc.). As done in the paper for the free-information case, we consider an extended game composed of two stages. First, each agent of taste $t \in [0, 1]$ can choose the tastes of the remaining $n - 1$ agents in her group. Second, the information-collection game described below is played.

2.1 Information Collection

When information is costly, each agent's decision is comprised of two layers in the information-collection game. Each agent decides, first, whether or not to collect information. Then, upon deciding to collect information, she decides on the source of that information. In analogy with the free-information case, we assume that an agent who is indifferent between acquiring and not acquiring a signal invests in information, and an agent who is indifferent between an A- and a B-signal chooses an A-signal. We focus our analysis on pure strategy equilibria (pertaining to both layers). Therefore, in equilibrium, the agent knows the number of other agents in the group who acquire information, and the analysis of the second layer boils down to that performed for the free-information case. However, it is important to note that some information collection equilibria may involve agents not collecting information.

Given a group of agents of tastes $t_1 \geq ... \geq t_n$, an information-collection equilibrium is characterized by the profile of chosen sources $(x_1, ..., x_n)$, where $x_i \in \{A, B, \varnothing\}$ is the source chosen by agent i (and $\varnothing$ stands for agent i not acquiring information).

As it turns out, given our tie-breaking rule, a pure equilibrium exists. We assume that the most (utilitarian) efficient such equilibrium is selected. As it turns out, mirroring Lemma 1 in the paper, there is a simple characterization of an efficient equilibrium, which we provide in Lemma A1.

Lemma A1 (Existence) *For any group of n agents with tastes $t_1 \geq t_2 \geq ... \geq t_n$, there exists $\tau^A \in \{0, ..., n\}$ and $\tau^B \in \{1, ..., n+1\}$, $\tau^B > \tau^A$, such that all agents $i \leq \tau^A$ acquiring the A-signal, all agents $i \geq \tau^B$ acquiring the B-signal, and all other agents not acquiring information, constitutes an efficient Nash equilibrium of the information-collection game.*

Given a group of n agents with tastes $t_1 \geq t_2 \geq ... \geq t_n$, Lemma A1 allows us to concentrate on equilibria $(x_1, ..., x_n)$ identified by two thresholds $\tau^A \in \{0, ..., n\}$ and $\tau^B \in \{1, ..., n+1\}$ such that $x_1 = ... = x_{\tau^A} = A$ and $x_{\tau^B} = ... = x_n = B$ (in particular, if $\tau^A = 0$ all agents choose a A-signal, and if $\tau^B = n$, all agents choose an B-signal). In words, we focus on equilibria in which any agent choosing the source A cares more about the issue α than does any agent not choosing any source or choosing the source B. When c is sufficiently small, equilibria of the information-collection game are isomorphic to those of the free-information case described in the paper. In particular, any equilibrium of the information-collection game is efficient and the number of A and B equilibrium signals is determined uniquely.

In the proof of Lemma A1 we show that equilibrium outcomes are ranked according to the volume of information collected on each issue. That is, whenever more A-signals are collected in one equilibrium relative to another, more B-signals will be collected as well in that equilibrium. As a consequence, all agents agree on their most preferred equilibrium outcome. Furthermore, the volume of information collected on each issue corresponding to an efficient equilibrium outcome is uniquely determined.

It is important to note that information collection equilibria may involve agents not collecting information. The fact that information is costly introduces a *free rider problem*. Indeed, in any equilibrium in which $\tau^B > \tau^A + 1$, the agents i, $\tau^A < i < \tau^B$, do not have enough incentives to collect information on either issue. Of course, if $c = 0$, the free rider problem disappears and all agents acquire a signal in equilibrium (i.e., $\tau^B = \tau^A + 1$).

Since Lemma A1 guarantees that the number of A-signals is determined uniquely in the efficient equilibria of the information-collection game, the agent's optimization problem in the first stage of the 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. Finally, we define stability in the first stage of the extended game as in Section 2 of the paper–that is, a group $(t_1, .., t_n)$ is stable if it is optimal for all its members–i.e., $(t_1, .., t_n) \in \cap_{i=1}^{n} O(t_i)$.

3 Group Composition

The goal of this section is to analyze the group properties entailed by the stability notion introduced in Section 2 of the paper. To do so, we first fix the taste parameter of one agent and identify that agent's optimal peer group choice. We then provide a full characterization of the stable groups. The following definitions are useful for our analysis.

First, denote by $n^A(t)$ the optimal number of A-signals the agent with taste parameter t would choose out of a total of n available signals. 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 $n^B(t) \equiv n - n^A(t)$. In words, $(n^A(t), n^B(t))$ represents the **unconstrained optimal allocation** of n signals for an agent of taste t. Naturally, $n^A(t)$ increases with t and with the group size n.

Second, for $x = A, B$, let $k_c^x(t)$ denote the maximal number of x-signals acquired in a group for which an agent of taste t is willing to acquire an x-signal rather than no signal at all. We term $k_c^A(t)$ the **attainable number of A-signals** for an individual of taste t. Formally, for any $c > 0$, $k_c^A(t)$ is the maximal integer h

such that, no matter how many w B-signals are acquired,

$$U(t,h,w) - U(t,h-1,w) = \frac{t}{2}\left(1-q_A\right)^{h-1} q_A \geq c. \tag{2}$$

Similarly, the **attainable number of** B**-signals,** $k_c^B(t)$, is defined for any $c > 0$ as the maximal integer h such that, no matter how many w A-signals are acquired,

$$U(t,w,h) - U(t,w,h-1) = \frac{1-t}{2}\left(1-q_B\right)^{h-1} q_B \geq c. \tag{3}$$

When $c = 0$, we denote $k_c^A(t) = k_c^B(t) = \infty$, corresponding to each agent willing to collect a free signal regardless of the number of signals already available. For any $t \in [0,1]$, we define $k_c(t) \equiv k_c^A(t) + k_c^B(t)$ as the **total attainable number of signals for an agent of taste** t.

Next, let $k_{\max}^A$ be the **maximal attainable number of** A**-signals** that corresponds to agents with the most extreme taste parameter $t = 1$. That is, $k_{\max}^A \equiv k_c^A(1)$. Analogously, for B-signals, the **maximal attainable number of** B**-signals** is $k_{\max}^B \equiv k_c^B(0)$. Note that $k_c^A(t)$ is increasing in t and $k_c^B(t)$ is decreasing in t. Therefore, $k_{\max}^A$ and $k_{\max}^B$ are the maximal number of A- and B-signals that can be expected to be acquired, respectively, in any equilibrium of the information-collection game.

Finally, let an α**-extremist** be an agent of taste t such that $k_c^A(t) = k_{\max}^A$. In words, α-extremists are agents who are willing to acquire the maximal number of A-signals. Likewise, β**-extremists** are agents of taste t such that $k_c^B(t) = k_{\max}^B$. It is easy to see that α- and β-extremists are agents with "sufficiently extreme" tastes to be willing to acquire the maximal possible number of signals, and their tastes lie in two intervals of the forms $\left[\underline{t}^A, 1\right]$ and $\left[0, \bar{t}^B\right]$, respectively.

3.1 Optimal Groups

Given a group size n, we now analyze the optimal group composition from the point of view of an agent with taste parameter $t \in [0,1]$.

As discussed in the paper, when information is free, any first-best group for the agent with taste parameter t is composed so that $n^A(t)$ agents collect an A-signal, and $n^B(t)$ agents collect a B-signal, 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, since extreme agents of taste $t = 1$ (or $t = 0$) always best respond with the choice of A- (or B-) signals, 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. When information is costly ($c > 0$), the pressure to choose more extreme individuals is even more pronounced as these are the agents with the highest willingness to acquire information. In other words, the presence of the free rider problem may induce an agent to select more polarized sets of peers.

When information is costly (i.e., $c > 0$), in choosing a group, an individual has to consider her peers' incentives to acquire information (not only their choice of information source). In particular, the number of x-signals cannot exceed $k^x_{\max}$ for any source $x = A, B$. Therefore, the agent of taste parameter t has hope of achieving her unconstrained optimum only if $n^x(t) \leq k^x_{\max}$ for $x = A, B$. The following Proposition describes the optimal groups.

Proposition A1 (Optimal Groups) *Consider an agent of taste parameter t. In any optimal group, the agent*

1. *achieves the unconstrained optimal allocation $(n^A(t), n^B(t))$ if and only if it is feasible, i.e., $n^x(t) \leq k^x_{\max}$ for $x = A, B$, and it induces the agent to invest in information, i.e., $n^x(t) \leq k^x_c(t)$ for at least one $x \in \{A, B\}$;*
2. *implements precisely $k^x_{\max}$ signals x for each $x = A, B$ for which $n^x(t) > k^x_{\max}$.*

3.2 Stable Groups

In the previous section we described the optimal group choice for any individual of a given taste t. However, in most applications, *all* members of a group of peers have some leverage in choosing the other members. The goal of this section is to expand the analysis of Section 3 of the paper to arbitrary information costs, showing how the structure of stable groups changes as the information gathering cost c increases.

First of all, note that *stable groups always exist*. Indeed, having n agents of taste $t = 1$ and $\min\{n, k^A_{\max}\}$ all acquiring an A-signal, or analogously, having n agents of taste $t = 0$ and $\min\{n, k^B_{\max}\}$ all acquiring a B-signal, both constitute stable groups. Our goal is to characterize all stable groups in the presence of information costs. We show that the structure of stable groups depends crucially on the group's size n. We avoid issues of equilibrium selection in that we assume that when a subgroup of agents of identical tastes acquires information, they cannot guarantee lower investment in information by shifting to a different group in which a different agent acquires their specific signals.[2]

While we present the analysis for different group sizes, our results have a natural analogue in terms of information gathering costs: increasing the size of the group is tantamount to increasing costs in that both make the free rider problem more severe. Formally, note that by increasing c, we reduce the number of attainable signals $k_c^x(t)$ for all t, $x = A, B$ (in particular, we lower $k_{\max}^A$ and $k_{\max}^B$). Thus, our results shed light on the effects of new technologies that decrease costs of information gathering (e.g., the introduction of the telephone, Internet, search engines, etc.). Similarly, they tie to periods of time in which media spread was experienced, effectively decreasing the costs of information acquisition.

Large Groups. Given the definition of $k_{\max}^A$ and $k_{\max}^B$, the maximal number of attainable signals that can conceivably be acquired in equilibrium is $k_{\max}^A + k_{\max}^B$. In that respect, we refer to groups containing more than $k_{\max}^A + k_{\max}^B$ members as *large*. The following proposition provides the characterization of large stable groups.

Proposition A2 (Stability – Large n) *When $n > k_{\max}^A + k_{\max}^B$, stable groups take one of the following forms:*

1. $k_{\max}^A$ α*-extremists and* $k_{\max}^B$ β*-extremists*;
2. n *agents with taste* $t = 1$*, or* n *agents with taste* $t = 0$*.*

The first part of Proposition A2 describes stable groups in which the number of signals gathered in the information-collection phase is maximized. When groups are very large, any agent with $t \in (0, 1)$ would desire a group in which the maximal amount of information on both issues is acquired. The only way to achieve this volume of information is to have a group in which at least $k_{\max}^A$ agents are α-extremists and at least $k_{\max}^B$ agents are β-extremists. Note that, while in the free-information case stability always entailed some degree of similarity, the first part of Proposition A2 suggests that for large group sizes, stability is consistent with *extreme group polarization*.[3] In the second part of Proposition A2 we describe groups that are *always*

[2] This assumption is important for the characterization of stable groups when information costs are high, and free riding emerge in stable groups. Naturally, if individuals assigned to acquire information could switch to an identical group and have their role as information gatherer or free rider determined randomly, more restrictions would be imposed on stable groups, and stable groups entailing free riders would cease to be stable.

[3] We refer to groups that contain at least one α-extremist and one β-extremist as *extremely polarized*. In fact, Proposition A2 suggests that as c decreases there will be a greater volume of agents on each side of the taste spectrum.

stable, in which all agents have the same, extreme, taste parameter. This is a knife-edge case in which all agents get *no* utility from signals on the issue they do not care about. In such groups, the maximal attainable number of signals is collected on the issue all of the members do care about.

Going back to the analogy between large group size n and high information cost c, as noted, a higher information cost c lowers both $k^A_{\max}$ and $k^B_{\max}$. Thus, Proposition A2 suggests that high costs induce extreme polarization.

Small Groups. In general, the optimal composition of signals for each agent entails the collection of signals from both sources. When groups are small, the free rider problem is weaker, implying the possibility of the unconstrained optimal allocation being consistent with stability. The following Lemma will be useful in linking stable groups with costly information to those identified in the free-information case through Proposition 4 in the paper. Recall that for any $t \in [0,1]$, we define $k_c(t) \equiv k^A_c(t) + k^B_c(t)$ as the **total attainable number of signals for an agent of taste** t. It is the maximal number of signals that a group of agents of the same taste t would be willing to acquire in equilibrium.

Lemma A2 (Individual Incentives and Group Size) *Whenever* $n \leq k_c(t)$, $n^x(t) \leq k^x_c(t)$ *for* $x = A, B$.

Lemma A2 links the size of the group and the personal incentives to acquire information of an individual of taste t. As long as group size is sufficiently small (relative to the total attainable number of signals for an agent of taste t), the agent engages in information acquisition when her unconstrained optimal allocation of signals is instated. To see why this is the case, suppose that $n^A(t)+n^B(t) = n \leq k_c(t)$ but, for instance, $n^A(t) > k^A_c(t)$. Then, it must be the case that $n^B(t) < k^B_c(t)$. Since $n^A(t)$ and $n^B(t)$ represent the unconstrained optimal allocation, they are selected in a way that (approximately) equates the marginal returns from signals on either source. However, since $n^B(t) < k^B_c(t)$, the marginal benefit from the $n^B(t)$-th B-signal is greater than c. Thus, the marginal benefit from the $n^A(t)$-th A-signal should be greater than c as well, in contradiction to $n^A(t) > k^A_c(t)$.

We now turn to the characterization of stable groups when $n < k^A_{\max} + k^B_{\max}$. First, some of the homogeneous groups described in Proposition 5 of the paper are still stable when information is costly. Denote by $m^A_c(t)$ and $m^B_c(t)$ the real numbers achieving equality within the constraints (2) and (3), respectively. For

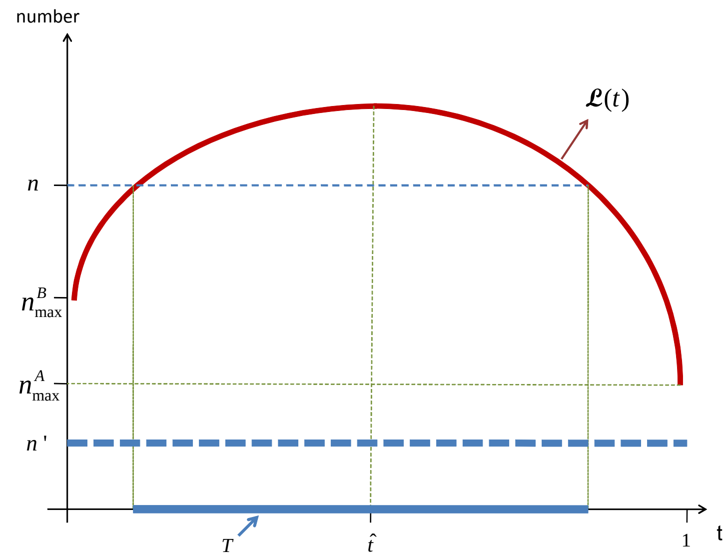

Figure 1: Small Group Size (Low Costs)

$x = A, B$, $k_c^x(t) = \lfloor m_c^x(t) \rfloor$ if $m_c^x(t) \geq 0$ and $k_c^x(t) = 0$ if $m_c^x(t) < 0$ (by construction, $m_c^x(t) < k_{\max}^x + 1$ for $x = A, B$). The number $m_c^x(t)$ captures the point at which an agent of taste t equates the marginal return from an x-signal to the cost c. The total attainable number of signals for an agent of taste t, or $k_c(t)$, is then approximated by the function $\mathcal{L}(t) = \max\left\{m_c^A(t), 0\right\} + \max\left\{m_c^B(t), 0\right\}$ depicted in Figure 1, which reaches its maximum at $\widehat{t} \in [0, 1]$.[4]

Looking at Figure 1, when $n \leq n_c(\hat{t})$, there is always a set of taste parameters T such that for any $t \in T$, $n \leq k_c(t)$. By Lemma A2, this guarantees that groups formed by agents sharing the same unconstrained optimal allocation (corresponding to stable groups in the free-information case) and with tastes within T, are able to achieve their optimal allocation in the costly-information case as well. That is, suppose stable groups of size n in the free-information case are characterized by the partition $\{T_k^n\}_{k=0}^n$. When information comes at a cost of c, for all the elements in the partition such that $T_k^n \cap T \neq \varnothing$, any group of n agents in $T_k^n \cap T$ is stable.

We now focus our attention on polarized groups. The extreme polarization appearing for large groups (Proposition A2) is not stable for smaller group sizes.[5] Nonetheless, for intermediate group sizes, ones for

[4]The concavity portrayed in Figure 1 corresponds to sufficiently low costs. As costs increase, extreme agents are willing to acquire information only on the issue they care most about, and $\mathcal{L}(t)$ becomes piece-wise concave.

[5]To see why, consider a polarized group in which some agents are α-extremists (and their optimal groups entail $k_{\max}^A$ A-signals),

which information on at least one dimension can be maximized (i.e., $n > \min\left\{k^A_{\max}, k^B_{\max}\right\}$), ***mildly polarized*** groups in which some agents have extreme tastes, while others have moderate (and similar) tastes, can be stable. Indeed, suppose that $n > k^A_{\max}$. We now illustrate that a stable group can be comprised of a sub-group of α-extremists that get $k^A_{\max}$ A-signals, and a sub-group of moderates, who all acquire B-signals. Formally, assume that $k^A_{\max}$ agents care sufficiently about issue α so that: (i) they are α-extremists (i.e., they have strong enough incentives to collect $k^A_{\max}$ A-signals); and (ii) their unconstrained optimal allocation involves at least $k^A_{\max}$ A-signals (so that they do not prefer groups with greater B-signal acquisition). To capture these restrictions, we define for $x = A, B$:

$$W^x = \left\{t \mid k^x_c(t) = k^x_{\max} \text{ and } n^x(t) \geq k^x_{\max}\right\}.$$

It is easy to see that W^A and W^B are intervals of the form $W^A = [\underline{t}, 1]$ and $W^B = \left[0, \bar{t}\right]$. Suppose $k^A_{\max}$ of the n agents have taste parameters in the interval W^A. By construction, these agents are in an optimal group whenever the number of A- and B-signals collected is $k^A_{\max}$ and $n - k^A_{\max}$, respectively.

We choose the remaining $n - k^A_{\max}$ agents to satisfy two conditions: (i) they care enough about issue β to have sufficient incentives to collect $n - k^A_{\max}$ B-signals; and (ii) they care enough about issue α so that they would not prefer a group in which more than $n - k^A_{\max}$ B-signals are collected. Formally, for $x, y = A, B$, $x \neq y$, we define:

$$Z^x = \left\{t \mid k^x_c(t) \geq n^x(t) \text{ and } n^x(t) = n - k^y_{\max}\right\}.$$

Thus, if we select the remaining $n - k^A_{\max}$ agents to have taste parameters within Z^B, by construction, these agents will be in an optimal group. In particular, combined with the $k^A_{\max}$ agents with tastes in W^A, they form a stable group. Note that for any taste parameter t in Z^B, $n^B(t) = n - k^A_{\max} < k^B_{\max}$. Therefore, Z^B ***does not contain the extreme taste parameter*** $t = 0$. This implies that the stable groups we just constructed involve a *milder degree of polarization* with respect to the large-group-size case.

The following proposition summarizes our discussion and provides the full characterization of stable groups for small group sizes (where we continue using the notation of $\{T^n_k\}^n_{k=0}$ for the partition corresponding

and some agents are β-extremists (and their optimal groups entail $k^B_{\max}$ B-signals). Clearly, when $n < k^A_{\max} + k^B_{\max}$, not all agents can be in their optimal group.

to stable groups in the free-information case).[6]

Proposition A3 (Stability – Small n) *When $n < k^A_{\max} + k^B_{\max}$, stable groups take one of the following forms:*

1. *If there is an interval T such that $n \le k_c(t)$ for all $t \in T$, then, for any $k = 0, ..., n$ for which $T \cap T^n_k \neq \varnothing$, a **homogeneous group** comprised of any n agents of tastes in $T \cap T^n_k$.*
2. *If $n > k^x_{\max}$, a **mildly polarized group** comprised of $n^x_{\max}$ agents with tastes in W^x, and $n - k^x_{\max}$ agents with tastes in Z^y.*
3. *n agents of taste $t = 1$, or n agents of taste $t = 0$.*

Regarding the case $n = k^A_{\max} + k^B_{\max}$, whenever $k^x_{\max} > 1$ for $x = A, B$, the groups described in Proposition A2 are the only stable ones. However, if $k^x_{\max} = 1$ for some x, say A, the full characterization of the stable groups includes additional types of groups consisting of $k^B_{\max}$ agents on the $t = 0$ extreme and one moderate agent. Since the full characterization involves some minor technical subtleties without adding qualitative novelties, we refer the interested reader to Proposition A4 presented in the proofs (Section 4) below.[7]

Note that a consequence of our results is that there is a set of moderate taste parameters such that individuals with those tastes can only be part of stable groups that are either small (emulating the free-information environment) or very large (in which the moderate agents free ride on the extremists in the group, who collect all the information).

The analogy between small group size n and small information cost c suggests that, as c decreases, some polarization can still persist in stable groups (point (2) of Proposition A3), but it is milder than the extreme polarization emerging for high c. Eventually, for sufficiently low c, polarized groups disappear. Moreover, as

[6] Note that whenever $n < \min\left\{k^A_{\max}, k^B_{\max}\right\}$ part (3) of Proposition A3 is subsumed in part (1).

[7] It is interesting to consider the consequences of side payments. When groups are small, so that stability entails the acquisition of the same signal profile of the free-information case, side payments have no consequence. For large groups, side payments allow agents to share the cost of information and invest (albeit indirectly) in more than one signal. In that case, the availability of side payments generates more information acquisition. In particular, for large group sizes, the introduction of side payment will tend to generate more similarity in stable groups. To see why, consider a polarized group (i.e., a group formed *only* by extemists on both issues). In such a group, an extremist on issue α will have an incentive to pay an extremist on issue β to acquire an A-signal rather than a B-signal. However, the same goal could have been achieved at a lower cost by selecting another extremist on issue α as a peer in the first place. This suggests that extreme polarization still arises in groups that are sufficiently large (but the lower bound on group size for extreme polarization to arise is higher than in our setting).

c decreases, homogeneous groups identical to those identified as stable in the free-information case emerge as stable (point (1) of Proposition A3).

A message that comes out of our analysis is that when *information gathering becomes cheaper (as a result, for instance, of better information technologies), stable groups tend to become more homogeneous.* As mentioned in the Introduction, this insight is backed up by a large body of empirical work. For example, the introduction of the telephone made social affiliations depend far more on shared interests (e.g., Sproull and Kiesler (1991)). Similarly, the introduction of the Internet is associated with a significant increase in the similarity of academic coauthors (see Rosenblat and Mobius (2004) and references therein).[8]

Finally, we notice that all results of this appendix are robust to the case in which each agent can acquire up to any number $h \geq 1$ of signals, each at cost c. However, if we consider a model in which agents can individually contribute any number of signals, homogeneous groups composed of only moderate types are harder to sustain as stable. Indeed, agents would have an incentive to substitute moderate peers for extreme ones, who will tend to acquire more signals. On the other hand, the other kinds of groups characterized in Proposition A3 are robust to this extension.

4 Proofs

Proof of Lemma A1. Let $t_1 \geq ... \geq t_n$. Each agent has to decide whether to acquire an A-signal, a B-signal, or forgo information gathering.

To construct an efficient equilibrium in the information-collection game, let μ^A be the maximal integer h such that $\frac{t_h}{2}(1-q_A)^{h-1} q_A \geq c$ (this is inequality (2) in the text). Similarly, let μ^B be the minimal integer h such that $\frac{(1-t_h)}{2}(1-q_B)^{h-1} q_B \geq c$ (this is inequality (3) in the text). First, consider the case in which $\mu^A + 1 \geq \mu^B$, so that all agents could be induced to acquire information. We first construct an equilibrium entailing all agents acquiring information. We consider an equilibrium as proposed by the Lemma's claim, so that $\tau^A = \tau^B - 1 \equiv \tau^*$. Note that if an agent of taste t prefers getting an A-signal over a B-signal, so would

[8] It is interesting to note the technical connection between the analysis of this section and a setting in which the group size n is an object of choice (besides its composition), and an individual connection cost $d > 0$ is incurred by every member for each additional agent added to the group. The marginal value of an additional signal determines the optimal size n. Since the function $\mathcal{L}(t)$ approximates the maximal number of signals that an agent of type t would be willing to invest in, the same function can be used to determine the optimal group size for an individual of type t. This suggests that, absent information costs, but accounting for connection costs, more moderate groups will be larger and groups containing extremists will be smaller.

any agent of taste $t' > t$. Similarly, if an agent of taste t prefers getting a B-signal, so would any agent of taste $t' < t$. In such an equilibrium, the agent with the lowest taste parameter who chooses an A-signal is the agent with taste t_{τ^*}. From our tie-breaking rule, it follows that the threshold τ^* is determined as the maximal $\tau \in \{1, ..., n\}$ for which agent τ *weakly* prefers an A-signal over a B-signal, or for which

$$U(t_\tau, \tau, n - \tau) \geq U(t_\tau, \tau - 1, n - \tau + 1)$$

is satisfied. This inequality is constraint (1) for taste t_τ. If (1) is not satisfied for any agent in the group (i.e., $U(t_1, 0, n) > U(t_1, 1, n - 1)$), then $\tau^A = \tau^* = 0$ and $\tau^B = 1$ defines an equilibrium. In order to show that choosing $\tau^* = 0$ if (1) is not satisfied for any positive integer and τ^* as the maximal integer between 1 and n satisfying (1) otherwise defines an equilibrium all that remains to be shown is that incentives to acquire information are satisfied. Notice that for any agent $\tau \leq \tau^*$,

$$\begin{aligned} U(t_\tau, \tau, n - \tau) - U(t_\tau, \tau - 1, n - \tau + 1) &= [U(t_\tau, \tau, n - \tau) - U(t_\tau, \tau - 1, n - \tau)] \\ &\quad - [U(t_\tau, \tau - 1, n - \tau + 1) - U(t_\tau, \tau - 1, n - \tau)] \geq 0 \end{aligned}$$

and so, the incentives to acquire an A-signal are greater than those to acquire a B-signal. Similarly, for agents $\tau > \tau^*$, the incentives to acquire a B-signal are greater than those to acquire an A-signal. Since $\mu^A + 1 \geq \mu^B$, it follows that the identified profile constitutes an equilibrium.

Consider now the case in which $\mu^A + 1 < \mu^B$, and define $\tau^A \equiv \mu^A$ and $\tau^B \equiv \mu^B$. From our definitions of μ^A and μ^B, in order to illustrate that the suggested profile constitutes an equilibrium, all that remains to be shown is that an agent acquiring a signal $x = A, B$ does not prefer to acquire a signal $y \neq x$ when all other agents follow the profile. Indeed, suppose that $i \leq \tau^A < \tau^B$ and observe that

$$U(t_i, \tau^A, n - \tau^B + 1) - c \geq U(t_i, \tau^A - 1, n - \tau^B + 1) > U(t_i, \tau^A - 1, n - \tau^B + 2) - c,$$

where the first inequality follows from inequality (2), and the second from the fact that $\mu^A + 1 < \mu^B$. Thus, an agent of taste t_i does not profit from deviating to a choice of a B-signal instead of an A-signal. An analogous argument holds for $i \geq \mu^B > \mu^A$.

Suppose now that there are two equilibria, one of which entails k^A A-signal and k^B B-signals and one that entails $\tilde{k}^A$ A-signals and $\tilde{k}^B$ B-signals. We now show that either $k^A \leq \tilde{k}^A$ and $k^B \leq \tilde{k}^B$ or $k^A \geq \tilde{k}^A$ and $k^B \geq \tilde{k}^B$. Suppose, for instance, that $k^A > \tilde{k}^A$ and $k^B < \tilde{k}^B$. This implies that there is an agent with taste t_i

that in the first equilibrium acquires an A-signal, and in the second equilibrium acquires either no signal or a B-signal. However, notice that for any such t_i

$$U(t_i, k^A, k^B) - U(t_i, k^A - 1, k^B) \leq U(t_i, \tilde{k}^A + 1, \tilde{k}^B) - U(t_i, \tilde{k}^A, \tilde{k}^B) \text{ and}$$

$$U(t_i, k^A, k^B) - U(t_i, k^A - 1, k^B + 1) \leq U(t_i, \tilde{k}^A + 1, \tilde{k}^B - 1) - U(t_i, \tilde{k}^A, \tilde{k}^B),$$

in contradiction to t_i using a best response in both equilibria. Other cases are shown similarly.

Since the equilibrium identified above establishes the maximal volume of signals, it follows that it is also the most efficient. ■

Proof of Proposition A1. In order to show that the conditions in (1) are sufficient, we construct an optimal group as follows. If an agent is selected to be part of the group and is to collect, say, an A-signal, then she must have a taste parameter t' such that (i) she prefers to gather an A-signal rather than a B-signal, that is, $t' \geq l(t)$, as described in the text for the $c = 0$ case, and (ii) she has enough incentives to gather an A-signal rather than no signal, that is $k_c^x(t') \geq n^x(t)$. Thus, $n^x(t) \leq k_{\max}^x$ is a necessary condition for the unconstrained optimal allocation $(n^A(t), n^B(t))$ to be achievable. Moreover, to achieve her optimal allocation $(n^A(t), n^B(t))$, the agent of taste parameter t has to have incentives *herself* to acquire a signal from at least one source. That is, $n^x(t) \leq k_c^x(t)$ for at least one $x \in \{A, B\}$. Note that if such incentives cannot be provided, meaning $n^x(t) > k_c^x(t)$ for $x = A, B$, an optimal group would entail the optimal allocation short of one signal.

To show (2), suppose, for example, that $n^A(t) > k_{\max}^A$ and $n^B(t) \leq k_{\max}^B$. Then, there is no selection of group members that allows our agent to achieve $n^A(t)$ signals from source A. Thus, after choosing $k_{\max}^A$ agents that collect A-signals (agents chosen in the interval $[\underline{t}^A, 1]$), the agent is better off choosing the remaining agents so that they collect as many B-signals as possible (this can be achieved, for instance, by choosing them in the interval $[0, \bar{t}^B]$). This could lead to more B-signals than in the unconstrained solution.

If both $n^A(t) > k_{\max}^A$ and $n^B(t) > k_{\max}^B$, the agent chooses a group in which $k_{\max}^A$ and $k_{\max}^B$ signals from sources A and B are collected, respectively. This can be achieved by selecting $k_{\max}^A$ agents in the interval $[\underline{t}^A, 1]$ and $k_{\max}^B$ in the interval $[0, \bar{t}^B]$. ■

Proof of Proposition A2. Suppose $n > k^A_{\max} + k^B_{\max}$. Any agent of taste $t = 0$ is in an optimal group as long as there are $k^A_{\max}$ agents who are acquiring an A-signal. Similarly, any agent of taste $t = 1$ is in an optimal group as long as there are $k^B_{\max}$ agents who are acquiring a B-signal. Any agent with $t \in (0, 1)$ is in an optimal group as long as there are $k^A_{\max}$ and $k^B_{\max}$ agents acquiring an A- and B-signal, respectively (indeed, she can contemplate a group with $k^A_{\max}$ and $k^B_{\max}$ agents of taste $t = 1$ and $t = 0$, respectively). Therefore, stable groups take one of the forms (1) or (2). ■

Proof of Lemma A2. Suppose that $n^A(t) > k^A_c(t)$. Then, it must be the case that $n^B(t) < k^B_c(t)$ (otherwise, $n = n^A(t) + n^B(t) > k^A_c(t) + k^B_c(t) = k_c(t)$, contrary to our assumption). That is, $k^B_c(t) \geq n^B(t) + 1$. In particular,

$$\frac{1-t}{2}\left[1-(1-q_B)^{n^B(t)+1}\right] - \frac{1-t}{2}\left[1-(1-q_B)^{n^B(t)}\right] \geq c.$$

Simple manipulations of the definitions of $n^A(t)$ and $n^B(t)$ imply

$$\begin{aligned} & t\left[1-(1-q_A)^{n^A(t)}\right] - \left[1-(1-q_A)^{n^A(t)-1}\right] \\ \geq\ & (1-t)\left[1-(1-q_B)^{n^B(t)+1}\right] - (1-t)\left[1-(1-q_B)^{n^B(t)}\right] \geq c \end{aligned}$$

and $k^A_c(t) \geq n^A(t)$, which contradicts our hypothesis. Identical arguments follow if $n^B(t) > k^B_c(t)$. ■

Proposition A4 *If $n = k^A_{\max} + k^B_{\max}$, stable groups take one of the following forms:*

1. *$k^A_{\max}$ agents whose taste falls in $\left[\underline{t}^A, 1\right]$ and $k^B_{\max}$ agents whose taste falls in $[0, \bar{t}^B]$;*
2. *n agents of taste $t = 1$, or n agents of taste $t = 0$;*
3. *If $k^A_{\max} = 1$, then $k^B_{\max}$ agents of taste 0, and one agent of taste $t \in (0, 1)$, where t satisfies one of the following:*
   (a) *$k^A_c(t) = 0$ and $U(t, 0, k^B_{\max}) \geq U(t, 1, k^B_{\max} - 1)$; or*
   (b) *$n^A(t) = k^A_c(t) = 1$.*

   Similarly if $k^B_{\max} = 1$.

Proof of Proposition A4. The analysis in Proposition A2 carries through as long as $k^A_{\max}, k^B_{\max} > 1$, and the classes of group compositions in points (1) and (2) constitute all of the stable allocations. Regarding (3), suppose that $k^A_{\max} = 1$ (analogous constructions can be performed when $k^B_{\max} = 1$). (a) Assume that $k^A_c(t) = 0$. First, consider t for which $n^A(t) = k^A_c(t) = 0$. The group consisting of an agent of type t and $k^B_{\max}$ agents of taste 0 is stable, since the agent with non-extreme taste parameter t does not have enough incentives to get information on one issue, even when she is the first to acquire a signal relevant to it. Moreover, the remaining agents have an extreme taste parameter, so have no incentive to acquire a signal other than the one pertaining to the issue they care most about. In this case, we have

$$U(t, 0, k^B_{\max}) - U(t, 1, k^B_{\max} - 1) \geq U(t, 0, k^B_{\max} + 1) - U(t, 1, k^B_{\max}) \geq 0,$$

where the last inequality follows from $n^A(t) = 0$. Second, consider the case in which $k^A_c(t) = 0$ and $n^A(t) = 1$. The group is stable since the condition $U(t, 0, k^B_{\max}) \geq U(t, 1, k^B_{\max} - 1)$ assures that the utility the agent gets from the $k^B_{\max}$-th B-signal is higher (or equal) to the utility she would get from the first A-signal. (b) Suppose $n^A(t) = k^A_c(t) = 1$. In this case, the group formed by $k^B_{\max}$ agents of taste $t = 0$ and one agent of type t is stable as the agent of taste t implements her unconstrained optimal allocation. ■

5 References

Che, Y-K. and N. Kartik (2009), “Opinions as Incentives,” Journal of Political Economy, 117(5), 815-860.

Gerardi, D. and L. Yariv (2008), “Costly Expertise,” American Economic Review, Papers and Proceedings, 98(2), 187-193.

Rosenblat, T. S. and M. M. Mobius (2004), “Getting Closer or Drifting Apart,” Quarterly Journal of Economics, 119(3), 971-1009.

Sproull, L. and S. Kiesler (1991), Connections – New Ways of Working in the Networked Organization, MIT Press, Cambridge, MA.