UAB
Universitat Autònoma de Barcelona

Essays on Social Behavior, Decision-Making, and sCommunication: Insights from Network Theory and Behavioral Economics

Defne Mevsim

Essays on Social Behavior, Decision-Making, and Communication: Insights from Network Theory and Behavioral Economics

Doctoral Thesis

Author: Defne Mevsim[1]

Supervisor: Prof. Pau Milán

A dissertation submitted to the Department of Economics of the Universitat Autònoma de Barcelona in partial fulfillment of the requirements for the degree of Doctor of Philosophy granted by the International Doctorate in Economic Analysis (IDEA) program.

September 2023

[1]I gratefully acknowledge financial support from the Spanish Ministry of Economy and Competitiveness through grant ECO2017-83534-P.

For every moment that brought me to this point.

–In memory of my sweet blue-eyed boy

Acknowledgments

I would like to express my deepest appreciation to my supervisor Pau Milán for his constant guidance, patience, and support throughout my master's and doctorate thesis. I would also like to extend my thanks to Ines Macho-Stadler for always being there when I needed support. I am deeply grateful to the faculty members from the Department of Economics and Economic History, especially Pau Olivella, and Jordi Masso for their helpful inputs to my research.

Words cannot express my gratitude to Faruk Yaşar for his invaluable support. I am certain that this endeavor would not have been possible without him.

Lastly, but definitely not least, I would like to thank my parents: my mother for being my greatest aspiration for pursuing a doctoral degree, and my father for providing unwavering support and ensuring that I had a genuine laugh each and every day.

Contents

List of Figures

Introduction

In an era characterized by intricate social networks and rapidly evolving information landscapes, understanding the mechanisms that underlie human behavior and decision-making is more critical than ever. This doctoral thesis serves as a comprehensive exploration into the realms of social behavior, decision-making, and communication. We present three interrelated chapters that collectively contribute to the existing literature by offering novel insights into how social networks influence individual choices and collective outcomes. These chapters, while distinct in their focus and methodology, are united in their aim to unravel the complexities of social interactions and their impact on individual and societal decisions.

In the first chapter, we introduce a novel framework to understand the dynamics of endogenous socialization, particularly focusing on high school friendships. We argue that the meeting process can be modeled as students' decisions about the costly activities they adopt during socialization. Our findings have significant implications for understanding minority group behavior, as we show that people from minority groups often coordinate on costly social activities, such as smoking, to meet others from their own group more frequently. This equilibrium is unique under various specifications of the cost structure, and we validate our model using Add Health data on high school students.

The second chapter takes a more applied approach by empirically analyzing a unique dataset from public family health centers in Izmir, Turkey. We focus on the effectiveness of various communication strategies employed by family care physicians in disseminating vaccine-related information. Our research categorizes these strategies into two primary classes: "broadcast" and "personalized." We find that personalized strategies, such as face-to-face consultations and phone calls, are more effective in educating individuals about COVID-19 infection and vaccines. However, we also note that increased information does not necessarily lead to higher vaccination rates, revealing the nuanced nature of information dissemination and its impact on public health decisions.

In the third chapter, we delve into the complexities of social learning within highly connected groups. Building on the existing literature, we scrutinize the conditions under which agents in a highly connected community become stuck in incorrect beliefs. Our findings indicate that the limitations of this 'stuckness' are closely tied to the model's assumptions about the simplicity of communication and the binary nature of states. We also extend the model to consider a discrete number of possible actions, providing a more generalized view of how communication coarseness affects social learning.

By synthesizing these chapters, we offer a multi-faceted view of the social and behavioral aspects that underlie decision-making processes. This thesis serves as a resource for policymakers, social scientists, and economists interested in leveraging social networks and communication strategies to influence public behavior and decision-making.

Chapter 1

Endogenous Socialization with Costly Behavior

Abstract

We build a simple framework of endogenous socialization that can account for the observed costly behavior patterns in high school friendships. We show that endogenizing the meeting process can be modeled as students' decisions about the costly activities they adopt during socialization. An equilibrium is derived in which people from minority groups coordinate on a costly social activity – such as smoking – and can meet with others from their own group more often than they would at random in the whole population. This equilibrium is unique under various specifications of the cost structure. We also empirically analyze friendship networks and costly behavior choices of high school students from Add Health data and show that people belonging to minority racial groups in their school adopt costly behavior more, controlling for other factors.

1.1 Introduction

Adolescence is commonly viewed as a transitional period between childhood and maturity when young people experience biological and psycho-social changes and engage in behaviors that have important implications for health risks and the adoption of healthy lifestyles (Millstein et al., 1994). There is an ongoing focus of public health literature on why adolescents generally adopt habits like smoking, drinking, and drug use, which are referred to as *risky behavior*. These are activities such that there is no uncertainty or risk in an economic sense so that we can think of them as *costly* behaviors. Many attempts in the literature exist to identify and explain the determinants of adolescent costly behaviors (Beyth-Marom et al., 1993).

A common explanation of adolescents' costly behavior inclinations is that teens suffer from peer pressure such that they try to engage in the same activity as their peers (Clark and Lohéac, 2007; Evans et al., 1992). That is one angle that assumes a predefined structure of friendships, and people are pressured to do what their friends do, given the social costs related to ostracism. However, in this paper, we take the opposite perspective, considering that the structure of friendships itself is endogenous, and how teenagers build it through costly activity. To be more accurate, it is not that people first form friendships and their friends force them to smoke; instead, people force themselves to smoke to form the "correct" type of friends. By correct, we mean that individuals have differential preferences over types of others'.

It is well-documented that people tend to form ties with those of a similar type, which is referred to as *homophily*. Homophily is a prevalent social life phenomenon concerning various dimensions such as race, gender, religion, age, profession, education, and socioeconomic status (McPherson et al., 2001).

Socialization patterns have been studied in network formation models that account for group identity and homophily (Bramoullé and Rogers, 2009; Bramoullé et al., 2012). The seminal paper of Currarini et al. (2009) develops a friendship formation model that investigates segregation patterns observed in social networks. Their paper shows that large groups exhibit positive homophily and small groups exhibit negative homophily. In other words, in their simple model, small groups can not find each other. But when people's observed socialization behaviors are analyzed empirically, there is overwhelming evidence that both the smaller and larger groups' friendships have homophily. To match this pattern, the authors are forced to assume that some "bias in the meetings" must be at work such that all groups can simultaneously meet their own type at faster rates than the ones implied by their population shares. However, they fail to provide any foundation for the bias in the meetings, as the parameter they define to capture the bias has no natural interpretation. Our research aims to endogenize the meeting process and give a general and natural characterization of the bias in the matching process, resulting in inbreeding homophily for both groups.

After biases in meeting processes have been shown to be a critical aspect of the existence of homophily in social networks, a recent paper by Currarini et al. (2016) provides a microfoundation of these biases in a discrete choice model. In their setting, agents have the option to inbreed, which means searching only their own types, or outbreed which means searching everyone in the population. They show a threshold equilibrium in which agents choose to outbreed if the total size of their group is above a threshold level, and they inbreed if it is below the threshold. Although they can endogenize meeting probabilities, their primary choice of "to inbreed" or "to outbreed" is not a natural economic/social choice that is to be determined by agents. Also, the conclusion that the equilibrium choice depends only on the absolute size of groups is an oversimplification of a rather pervasive and complicated phenomenon. Accordingly, their setting lacks the necessary realistic properties.

In this paper, we build an endogenous model of socialization that explains the observed costly behavior patterns in high-school friendships. In this complete information model, agents choose among two different social activities, which differ in cost. While doing that activity – such as smoking or drinking – they meet with others who also choose to do the same activity and form friendships with them. Each agent belongs to one of two distinct groups with potentially different sizes and gains utility from the friendships they form by meeting with each other uniformly *within* an activity.

We show an equilibrium in which agents from the minority group coordinate on a costly social activity, and by doing so, they get to meet with people of their own kind more often than they would do at random in the whole population. We also show that this equilibrium is a unique one under various specifications of the model. We also perform comparative statics and show that the time dedicated to costly activities increases as the group share decreases.

We also analyze the well-known Add Health high-school students' data, show the empirical patterns of homophily in students' racial groups, and analyze their costly behavior choices. We show that the time engaged in smoking/drinking responds to the population shares as predicted by the model. The data suggests that people belonging to minority racial groups in their school adopt costly behavior more on average relative to the school average, controlling for other factors. More specifically, we find that when comparing two students from some specific racial group, the student from a minority group adopts higher levels of costly behavior (i.e., smoking, drinking, and racing) than the one from a majority group. This result is in line with what the theoretical model suggests. It implies that policies that aim to tackle the costly behavior attitudes of racially isolated minority groups of high-school students also have a positive externality on racial segregation in the school.

The organization of the paper is as follows: The theoretical model is introduced in Section 1.2, we then present its results in Section 1.3. Consequently, in Section 1.4, we analyze the homophily and costly behavior in high-school friendships data and show the findings. Finally, we discuss and conclude in Section 2.5.

1.2 The Model

Here, we present the general environment of the theoretical model, describe the matching process that enables the socialization of agents, and mention how homophily of friendships is measured in this setting. Then we give a simple example of the simplest case of the model where agents socialize while doing an activity.

1.2.1 General Environment

There is a continuum of agents with two types, which can refer to ethnicity, gender, age, profession, or some deterministic and observable trait of people. The mass of type i agents in the population is indicated by N_i for $i \in \{1,2\}$ with $N_1 < N_2$, and the share of group i in the population is $w_i = \frac{N_i}{N_1+N_2}$.

Each agent chooses how much time to spend on two different activities, referred to as A and B, where we denote a generic activity by r. The time choices of agents are non-negative, and each of them is endowed with one unit of time. So the strategy of a type i agent is

$$(t_{iA}, t_{iB}) \in \mathbb{R}_+^2 \text{ and } t_{iA} + t_{iB} = 1.$$

Hence we can think of this setting as if individuals were partitioning their day between two activities, each with an associated cost.

Without loss of generality, it is assumed that activity A is a costly activity, such as smoking or drinking, and from now on activity A will simply be referred to as *smoking*.

Agents incur a convex health cost for each unit of time spent in the costly activity, $\frac{h}{2}t_{iA}^2$ where h is a parameter that governs the curvature of the cost. Therefore the additional harm they cause to their health with an extra unit of time spent smoking increases with the total amount of time they spend smoking.

While in a particular activity, all agents meet with others who also choose to do the same activity and form friendships with them uniformly at random. Agents gain utility from socializing with all types of individuals. The number of the same type friends of a type i agent is denoted by s_i, and the number of her other type friends is d_i. Notice that s and d are functions of the times the agent spends doing each activity so we denote them as $s(t_{iA}, t_{iB})$ and $d(t_{iA}, t_{iB})$.

The utility of agents is assumed to be a linear function of the number of her friends, $s(t_{iA}, t_{iB})$ and $d(t_{iA}, t_{iB})$, so the total payoff of a type i agent is the following:

$$U(s(t_{iA}, t_{iB}), d(t_{iA}, t_{iB})) = s(t_{iA}, t_{iB}) + \gamma d(t_{iA}, t_{iB}) - \frac{h}{2}t_{iA}^2.$$

Here, the parameter $\gamma \in (0, 1)$ captures the relative bias in preferences between same and other type friendships, where it is assumed that an agent experiences higher marginal benefits from same type friendships than from other type friendships, i.e. $\gamma < 1$.

1.2.2 Matching Process

While spending time on a specific activity, agents are assumed to meet with others who are also doing the same activity in a continuous matching process.

There is a new inflow of type i agents per unit of time, which is N_i. Since doing an activity can be thought of as spending time in a particular room with others who are also choosing that activity, the actual stock of type i agents in Room r is $t_{ir}N_i$. Accordingly, a particular type can increase the stock of their people in a room either by increasing the time they spend in the matching process of Room r or by increasing their flow.

Mathematically, the continuous matching process in Room r can be defined by a 2×2 matrix $q^r \in [0,1] \times [0,1]$ with q^r_{ij} component signifying the meeting probability that type i meets a type j agent per unit of time:

$$q^r = F(t_{1r}N_1, t_{2r}N_2).$$

We assume that agents meet each other uniformly at random within a certain room. We call this an *unbiased matching process*. For a particular Room r and distinct types i and j, the meeting probability that a type i agent meets with someone from her own type is

$$q_{ir} = \frac{t_{ir}N_i}{t_{ir}N_i + t_{jr}N_j}.$$

Accordingly, a type i agent, who spends t_{ir} amount of time in Room r, forms a total of t_{ir} friendships. From this, proportion $q_{ir}t_{ir}$ is with her own type friends, whereas proportion $(1-q_{ir})t_{ir}$ is with the other type. This follows from a mean-field approximation (Currarini et al., 2009).

Following the mean-field approximation, the total mass of the same type and other type friends of a type i agent, s_i and d_i, can be expressed in terms of the meeting probabilities within each activity as follows,

$$s(t_{iA}, t_{iB}) = t_{iA}q_{iA} + t_{iB}q_{iB},$$
$$d(t_{iA}, t_{iB}) = t_{iA}(1 - q_{iA}) + t_{iB}(1 - q_{iB}).$$

It means that, for instance, the mass of the same type of friends of a type i agent is equal to the sum of the terms $t_{ir}q_{ir}$ across all possible r activities. The term $t_{ir}q_{ir}$ is the expected mass of the same type of friends of a type-i agent when she chooses to spend t_{ir} time in Room r. By using the mean-field approximation, agents take the expected number of friends as their actual number of friends.

1.2.3 Measuring Homophily

In order to measure how homophilous friendships are in equilibrium, we define the following measures.

Homophily Index is defined by

$$H_i = \frac{s_i}{s_i + d_i},$$

where $s_i = s(t_{iA}, t_{iB})$ and $d_i = d(t_{iA}, t_{iB})$.

Homophily index H_i simply measures the fraction of type i agents' friendships with their own kind.

In the equilibrium, if this fraction of friendships for type i is higher than the population share of type i, that is w_i, it means that type i agents meet more often than they would normally meet uniformly random in the whole population. In this case, we say that there is *positive homophily* for type i. So, if $H_i > w_i$, then type i has positive homophily. Conversely, if $H_i < w_i$, then it means that the fraction of friendships is lower than the population share, and type i agents are meeting with each other less frequently than they would meet uniformly random, i.e., type i has *negative homophily*.

Even when every agent meets uniformly random in the whole population, it is unsurprising that H_i is higher for bigger groups. Since H_i measures the fraction of same-type friendships for type i, if everyone meets at random, they will meet with the large group members more often than the small group as the population share of the large group is higher by definition. Hence another measure of homophily is needed to see how biased the friendships are toward their own types, beyond the effect of population shares. That is what the following measure captures.

Inbreeding Homophily Index is

$$IH_i = \frac{H_i - w_i}{1 - w_i}.$$

It measures the amount of bias in friendships with respect to the groups' "baseline" homophily level that would be expected under a uniform random meeting that reflected groups' population shares.

1.2.4 Simple Example: One Activity

Here we mention one benchmark in this setting where it is assumed that there is only one activity. This is the case that Currarini et al. (2009) focus on. We show the calculations in Appendix A.1.

In this case, there are no different activities that agents choose among; in other words, there is only one room in which they socialize with each other. So a type i agent now only chooses t_i, how much time to spend socializing. For each unit of time spent, they pay a standard constant cost of socialization $c > 0$.

Results show that the large group socializes more than the small group. It is not surprising since there is a bias in preferences toward the same type friendships, and since the large group has more people, they find socializing more attractive and choose to spend more time socializing.

Consequently, the stock of the large group in the meeting process is much higher than the stock of the small group, which makes it much more likely to be matched with a large group member in the meeting process. Since within an activity, the matching is unbiased, it means that it is also very likely - more than what their population shares would lead to - for a small group member to be matched with someone from the large group. Therefore, in the equilibrium, we have positive homophily for the large group and negative homophily for the small group (shown in Appendix A.1).

1.3 Results

In this section, we first define the solution concept of the model, and then state a proposition showing the equilibria of the model. Then, we compare the three equilibria by Pareto-ranking them. Lastly, we present several modifications to the cost structure and show the unique equilibrium of the model.

1.3.1 Steady-State Equilibrium

As a solution concept for the model, we adopt the concept of *Steady-State Equilibrium* from Currarini et al. (2009).

Definition 1.1 (Steady-State Equilibrium). *The profiles* (t_{1A}, t_{1B}) *and* (t_{2A}, t_{2B}) *are a steady-state equilibrium satisfying the following conditions:*

- t_{iA} *and* t_{iB} *solve the following optimization problem of agents of type* $i \in \{1, 2\}$*:*

$$\max_{t_{iA}, t_{iB}} t_{iA}.q_{iA} + t_{iB}.q_{iB} + \gamma.(t_{iA}.(1 - q_{iA}) + t_{iB}.(1 - q_{iB})) - \frac{h}{2} t_{iA}^2,$$

where $t_{iA} + t_{iB} = 1$ *for all* $i \in \{1, 2\}$*.*

- *Strategies of agents determine the meeting probabilities of the types (Unbiased matching process is assumed within rooms). For each* $r \in \{A, B\}$*:*

$$q_{1r} = \frac{t_{1r}N_1}{t_{1r}N_1 + t_{2r}N_2},$$
$$q_{2r} = 1 - q_{1r} = \frac{t_{2r}N_2}{t_{1r}N_1 + t_{2r}N_2}.$$

1.3.2 Equilibrium of the Model

Proposition 1.1. *Assume that* $N_2 > N_1$ *(type-1 is the less numerous group) and Room A is the costly activity room. If* $h > 1 - \gamma$ *(The curvature of the convex health cost is sufficiently high compared to the marginal change in utility for an increase in time devoted to the costly activity.), then the model has three steady-state equilibria:*

1. $t_{1A} = t' \in (0, 1)$ *and* $t_{2A} = 0$.
2. $t_{1A} = 0$ *and* $t_{2A} = t'' \in (0, 1)$.
3. $t_{1A} = 0$ *and* $t_{2A} = 0$.

where it always holds that:

$$t' > t''.$$

Proof. In Appendix A.2. □

This proposition shows an equilibrium in which the smaller group can differentiate themselves from the large group, and get to meet with their own kind more often, given that the marginal cost of costly activity is higher than the difference between the marginal utilities of having one more own type and different type friend.

Moreover, this proposition states the three equilibria of the model. More specifically, an equilibrium of the model exists – equilibrium (1) – such that the minority group spends a positive amount of time engaged in the costly activity, whereas the majority group only spends time engaged in the costless activity. In equilibrium (2), the majority group is the only one who spends a positive amount of time engaged in the costly activity, while the minority group devotes all of their time to the riskless activity. Additionally, in these two equilibria the outcome implies that $H_1 > w_1$, $H_2 > w_2$, i.e., there is positive homophily for both groups.

Given that the other type is not engaging in the costly activity at all, it is always optimal for a type to devote some positive amount of time to the costly activity, although it is costly. In this case, the probability of meeting with their own type while doing the costly activity is the highest it can be, which is 1. So, in equilibrium (1) and (2), one of the types chooses to spend some time engaged in the costly activity, and while doing that activity, they get to meet with people *only* from their own group.

Furthermore, this proposition says that the time that the minority group spends on the costly activity in equilibrium (1) is always more than the time that the majority group chooses to spend in equilibrium (2). The reason comes from the fact that in equilibrium (2), the probability of meeting with someone from the minority group while engaging in the non-costly activity is quite low, not only because the other group is spending all of their time in the non-costly activity, but also minority group is fewer in numbers. Hence, the minority group is always more willing to pay the extra cost of costly activity to differentiate themselves from others.

In Appendix A.2, we consider all the possible cases that can result in equilibrium and show that the three equilibria stated in Proposition 1 are the only equilibria of the model.

1.3.3 Comparison of the Equilibria

Proposition 1 says that the optimal amount of time that the minority group prefers to do the costly activity in the equilibrium where the minority is the only group doing the costly activity is higher than the optimal amount of time that the majority prefers to do the costly activity in the equilibrium where the majority is the only group doing the costly activity.

In the equilibrium where the minority is the only group that taking part in the costly activity, the optimality condition for $t_{1A} = t' \in (0,1)$ of type 1 implied by the First-Order Condition of her maximization problem is

$$(q_{1A} - q_{1B})(1-\gamma) = ht_{1A}.$$

Plugging in the meeting probabilities induced by the choices of agents, $q_{1A} = 1$ and $q_{1B} = \frac{(1-t')N_1}{(1-t')N_1+N_2}$, the condition becomes

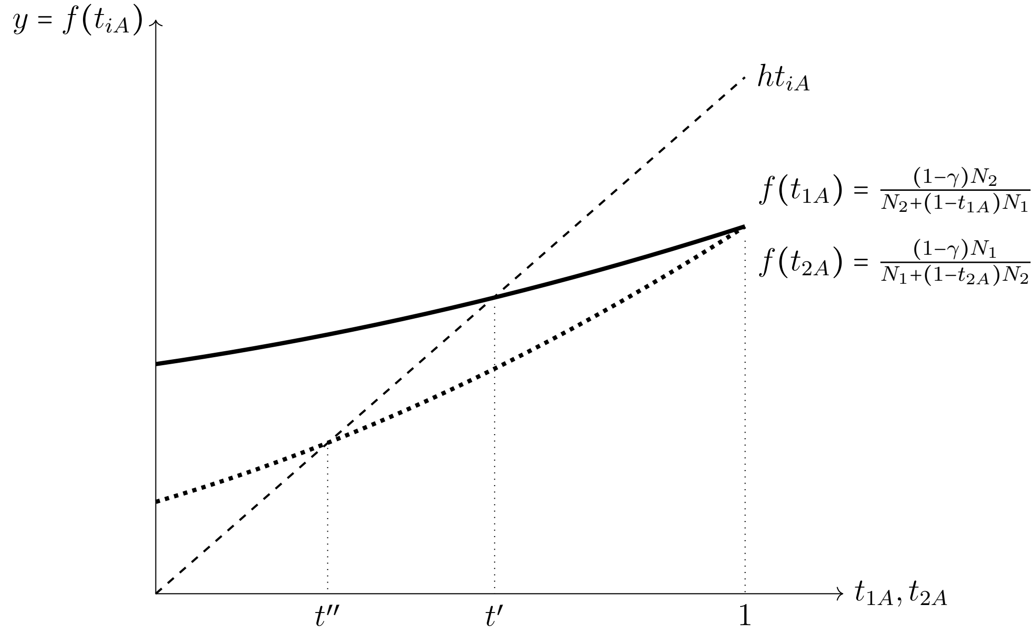
$$(1 - \frac{(1-t')N_1}{(1-t')N_1 + N_2})(1-\gamma) = ht'. \tag{1.1}$$

Symmetrically, in the other equilibrium where the majority is the only group taking part in the costly activity, the optimality condition for $t_{2A} = t'' \in (0,1)$ of type-2 is

$$(1 - \frac{(1-t'')N_2}{(1-t'')N_2 + N_1})(1-\gamma) = ht''. \tag{1.2}$$

The plot 1.1 depicts the comparison between the two optimal solutions t' and t'', fixing the values of parameters γ, N_1 and N_2.

Figure 1.1: General Optimality Condition



Here, the horizontal axis is the time spent in the costly activity for both types. The dashed line is the right-hand side of either of the equations (1.1) and (1.2), and it shows the marginal cost of increasing the time devoted to costly activity for possible levels of time choices. The solid curve is the left-hand side of equation (1.1), which shows the marginal gain in the utility of type 1 agents coming from choosing an incrementally longer time spent in the costly activity. The dotted curve is the left-hand side of equation (1.2), which demonstrates the similar marginal gain in utility for type 2.

Notice that the closed-form solution of t' is the solution for equation (1.1), and it is determined at the intersection of the dashed line and the solid curve. Similarly, the closed-form solution t'' of equation (1.2) is depicted at the intersection of the dashed line and the dotted curve.

Because of the fact that $N_2 > N_1$, i.e., type 1 agents are the minority group in the population, the value of t' always lies above the value of t''. In other words, the time devoted to the costly activity is always higher for the minority group than for the majority group, given that the other group is not engaging in the costly activity. As a consequence of the differences in the population shares of different types, the minority group always finds engaging in costly activity more profitable to socialize more frequently with members of their smaller group.

Comparative Statics

Here, we do comparative statics exercises in order to demonstrate how the endogenous variables of the model, t' and t'' change as a result of the changes in the parameters of the model: the bias in the preferences γ, and the mass of agents N_1 and N_2. The derivations are presented in Appendix A.3.

Time choices of both groups devoted to the costly activity, t' and t'', are decreasing in their own group's mass of agents and increasing in the other group's mass of agents. This result is not surprising as the main motive of agents for engaging in the costly activity is to differentiate themselves from the other group and improve the probability of meeting with someone from their own group. An increase in one's own group's mass of people leads to a rise in their meeting probability, making them reduce their costly activity. In contrast, an increase in other group's mass of people actually worsens their meeting probability, so they want to engage in the costly activity more than before.

More interestingly, this result also means that as the difference between N_2 and N_1 gets larger (either by an increasing N_2 or decreasing N_1), i.e., the size of the minority group gets smaller relative to the majority, the time choice of both groups will be more differentiated from each other, i.e., the difference between t' and t'' gets larger. Hence in a population where the group sizes are significantly differentiated from each other, we would expect to have two disparate time choices engaging in costly activity from the two types in the two equilibria where the costly activity choice of the minority group is always higher than the other.

As the preference bias of agents toward their own type increases – as γ decreases –, the choices of time devoted to costly activity for both types increase. Hence, now that people value their own type of friendship relatively much more than their friendships with the other type, they are more willing to pay the extra cost and engage in costly activity more.

Lastly, as h, which is the curvature of the convex health cost, increases, time choices of both types decrease. Agents choose a lower level of time to engage in the costly activity now that each positive level of time devoted to the costly activity costs more.

1.3.4 Pareto-Ranking of Equilibria

Proposition 1.2. *The equilibrium where only the minority engages in the costly activity (equilibrium (1)) strictly Pareto Dominates both of the equilibria where only the majority engages in the costly activity (equilibrium (2)) and no one engages in the costly activity (equilibrium (3)) if and only if* $2N_1 < N_2$.

Proof. In Appendix A.4. □

This proposition states that if the population is extensively more populated by the majority group relative to the minority group, both types are strictly better off in the equilibrium where only the minority engages in costly activity, than the other two possible equilibria.

The fact that the majority group is better off in the equilibrium where the minority group is the only group engaging in the costly activity is not surprising since the majority does not have to pay the extra health cost, and all of the members of their crowded group are doing the same activity while socializing with everyone from their own group who they get marginally higher utility from socializing with.

On the other hand, for the minority group, their meeting probability in the equilibrium where only the majority engages in costly activity is very low, not only because the minority has a lower number of people in general but also because the majority is spending a greater amount of time engaged in the non-costly activity, i.e., $1 - t' < 1 - t''$ since $t' > t''$. This would mean that even if the minority pays the extra health cost in the equilibrium where only the minority engages in costly activity, they are better off given that $2N_1 < N_2$, the minority is a substantially smaller group of people relative to the majority.

1.3.5 Extension: Modifications on the Cost Structure

In this section, we consider two different modifications to the cost structure of the model and show that the equilibrium where the small group is engaging in the costly activity is the unique equilibrium under several specifications of the cost function.

The common property of the specified cost function modifications is that they both have a discontinuous jump from $t_{iA} = 0$ to some small $t_{iA} = \epsilon > 0$, and they are continuous, strictly increasing, and convex for the interval $t_{iA} \in (0, 1)$. Given that the cost function holds these properties, the equilibrium where the minority is engaging in the costly activity can be sustained as the unique equilibrium with some restrictions on the parameters. Here, we present two such examples.

Fixed Cost

Here, consider the case where the total cost of activity A for any $t_{iA} > 0$ is the following:

$$\text{Total Cost } (t_{iA}) = R + \frac{h}{2} t_{iA}^2$$

where $R > 0$ is the *fixed cost* of the costly activity. This fixed cost is a lump sum cost of the activity that the individual must pay for the first marginal unit of time devoted to the costly activity. In the context of smoking and teenagers, it can be interpreted as the social cost of smoking that the teenager has to bear, coming from the guilt that she feels when she acts against her family's wishes.

Proposition 1.3. *If $h > 1 - \gamma$ and R satisfies*

$$\frac{N_1}{N_2} G(N_1, N_2, h, \gamma) < R < \frac{N_2}{N_1} G(N_1, N_2, h, \gamma)$$

where $G(N_1, N_2, h, \gamma) > 0$ is a function of N_1, N_2, h and γ, the model has two equilibria:

- $t^*_{1A} \in (0, 1)$ *and* $t^*_{2A} = 0$.
- $t^*_{1A} = 0$ *and* $t^*_{2A} = 0$.

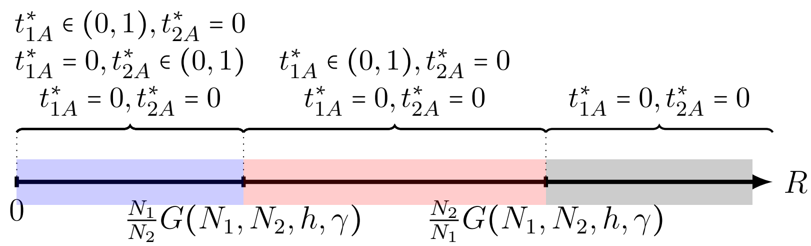
Proof. In Appendix A.5. □

Proposition 3 shows that if the cost function has a fixed cost component as an addition to the convex part, there exists an interval for fixed cost R where only the equilibrium where the minority is engaging in the costly activity survives, aside from the pathological equilibrium where no one is engaging in the costly activity.

The intuition is that given the additional cost is high enough, the only type that accepts to pay the cost is the minority. Because in the case where they do not pay this cost, they would face very low meeting probabilities while doing the costless activity B. Hence, their threshold of not accepting to pay for the costly activity is always higher than the threshold of more numerous type 2.

Moreover, conditional on all possible values of R, the equilibria of the model are shown in Figure 1.2.

Figure 1.2: Depiction of the Existence of Equilibria



In general, given that R is sufficiently high, i.e., $R > \frac{N_2}{N_1}G(N_1, N_2, h, \gamma)$, no one agrees to pay for the costly activity, so only the equilibrium where no one is devoting time to activity A exists. Also, for sufficiently low R, i.e., $R < \frac{N_1}{N_2}G(N_1, N_2, h, \gamma)$, all three of the original model's equilibria survive because both types are willing to pay the fixed cost of activity A. But, a possible moderate level of R exists, where the equilibrium where the minority is engaging in the costly activity exists and the equilibrium where the majority is engaging in the costly activity does not exist anymore.

Moreover, as the population share of the small group decreases (by decreasing N_1 or increasing N_2), the interval of R for which the only type spending time in costly activity is the small group - the area in between- increases.

Linear Quadratic Cost

An alternative way to obtain the same result is to let the total cost of activity A be, for any $t_{iA} > 0$,

$$\text{Total Cost } (t_{iA}) = rt_{iA} + \frac{h}{2}t_{iA}^2$$

where $r, h > 0$ are parameters of the cost function.

Proposition 1.4. *If* $h + r > 1 - \gamma$, *and* $\frac{N_1}{N_2} < \frac{(1-\gamma)-r}{r} < \frac{N_2}{N_1}$, *the model has two equilibria:*

- $t^*_{1A} \in (0,1)$ *and* $t^*_{2A} = 0$.

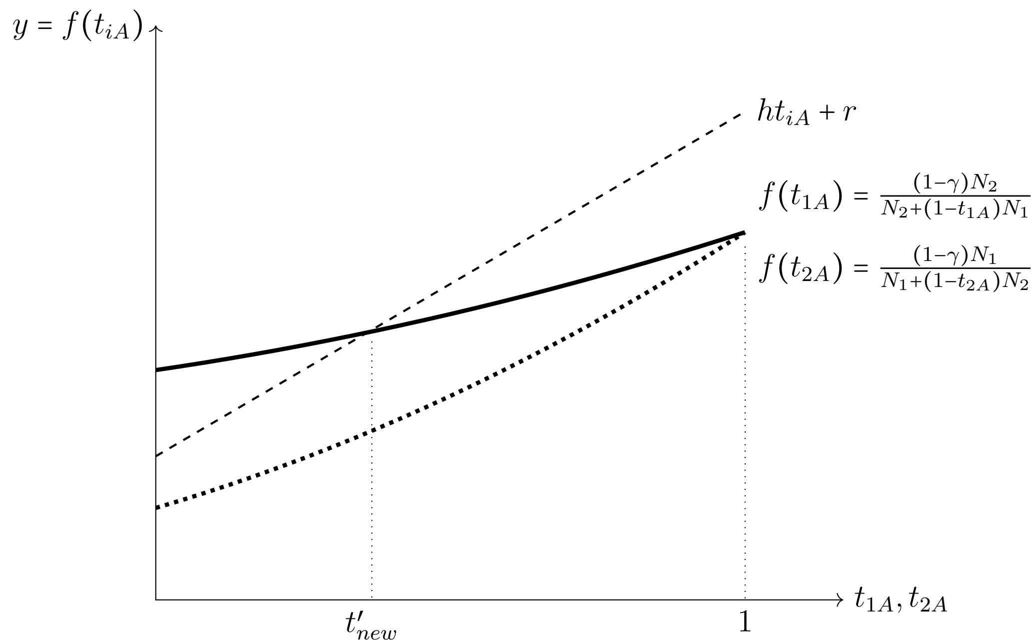
- $t^*_{2A} = 0$ *and* $t^*_{1A} = 0$.

Proof. In Appendix A.6. □

Proposition 4 shows that if the cost function is a convex function with this specific form, such that the discontinuity of the marginal cost function at 0, which is governed by r, lies within an interval depending on the population shares, only the equilibrium where the minority is engaging in the costly activity survives, aside from the pathological equilibrium where no one is engaging in the costly activity.

In this case, the optimality condition is shown in Figure 1.3.

Figure 1.3: Optimality Condition - Linear Quadratic Cost



This plot is very similar to the Figure 1.1. The only difference is the marginal cost function's upward shift by r. If the shift amount is high enough that the marginal cost line does not cross the marginal benefit curve of type 2 within the $(0, 1)$ interval, there is no equilibrium in which the majority is engaging in the costly activity.

Since the marginal benefit curve of the small group lies above the marginal benefit curve of the large group, we can always find a range for r such that there exists an equilibrium where the minority is engaging in the costly activity, which is now depicted by the time level t'_{new}.

1.4 Empirical Analysis

1.4.1 Data

This paper uses the data from Wave I of the National Longitudinal Study of Adolescent Health (Add Health). It is a longitudinal, school-based adolescent survey collected over several years since 1994. Using an implicit stratification procedure (stratified by region, urbanicity, ethnic mix, and size), researchers selected a nationally representative sample of private and public high schools in the United States, alongside their largest feeder school, typically middle school or junior high. This selection procedure resulted in a national sample of 142 schools for which network information as well as demographic and behavioral data exists.

The network data for each school is based on the reports of friendships by each student. In total, 90118 students provided global network information for each school. Each student received a list of students in their own school. Students were permitted to identify up to five female and five male friends, in total ten friends, from this roster. Although the maximum number of nominations is limited to ten, this constraint affects a few students. In the total data set, only 3% of students nominate ten in-school friends, 23% nominate five females, and 24% nominate five male friends. Since friendship nominations were recorded by student identification numbers located in rosters, it is possible to link together most students in schools to create a social network. In the network we constructed from this data, a tie between two subjects is present if either student nominates the other as a friend.

The ethnicity/race data is a self-defined choice by respondents from a listing of five categories – white, African-American, Asian, Native American, and other – alongside another question identifying those with Hispanic origin.

The costly behavior variables available in the data are questions about smoking, drinking alcohol, getting drunk, and, racing. They are answers to the questions "During the past 12 months, how often did you smoke cigarettes?", "... drink beer, wine, or liquor?", "... get drunk?", "... race on a bike, skateboard, or car?". The answers have seven categories varying as "never," "once or twice," "once a month or less," "2 or 3 days a month", "once or twice a week," "3 to 5 days a week" and "nearly every day"; aimed to capture the intensity of costly habits.

1.4.2 Patterns of Homophily in High-School Friendships

Here we introduce the results of the empirical analysis of friendship patterns of high school students using Add Health data.

Figure 1.4: Friendship Network in a US High-School from Add Health

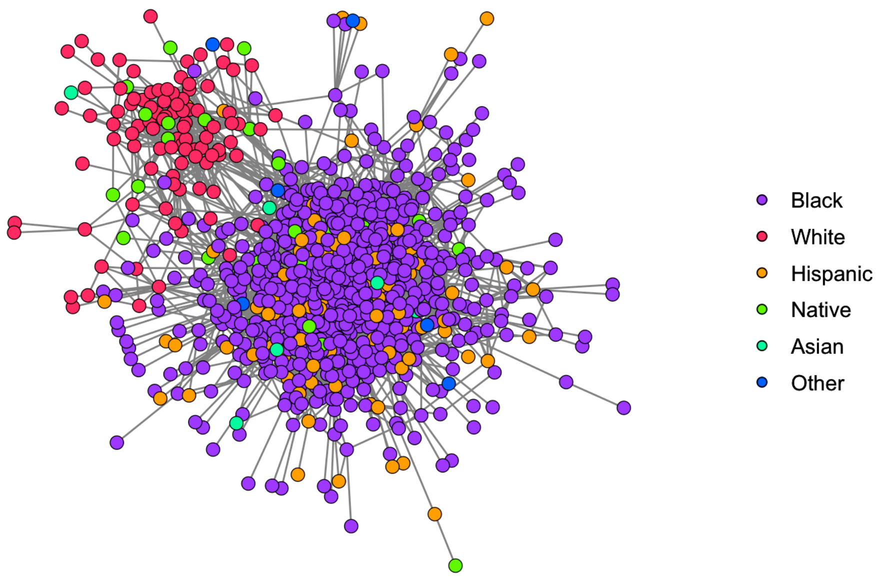


Figure 1.4 shows an example of a network of friendships constructed for one of the high schools from Add Health. Every node represents a particular student, and the node is colored based on the student's ethnicity. The size of nodes is adjusted based on their degree so that more popular students are shown with a larger node.

In this network, there is highly noticeable positive homophily in friendships, as the students from black and white groups form two distinct communities. Hence, their friendships have high levels of positive homophily. On the other hand, the other people from sporadic minority groups fail to form a distinct community, and they integrate more with other races, which implies that there is negative homophily for them.

To see the homophily patterns in the whole data set of 142 high schools with six different racial categories that the students can choose, we form a variable about the measure of homophily of a specific racial group in their corresponding school. To do that, we look at the ratio of the number of a student's friends from her racial group to her total number of friends, following the definition of *homophily* in Section 1.2.3. As this ratio approaches from zero to one, her level of homophily increases. Then we construct a variable by taking an average of the homophily measure of each student belonging to a particular racial group in their corresponding school.

Figure 1.5: Patterns of Homophily by Population Share of Racial Groups

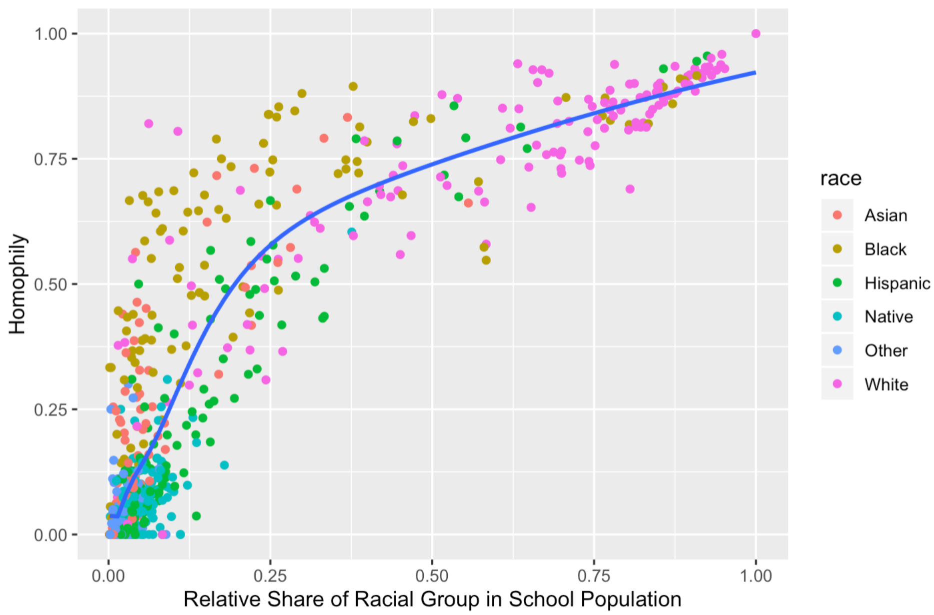


Table 1.1: Output for Regression of Homophily on Population Shares

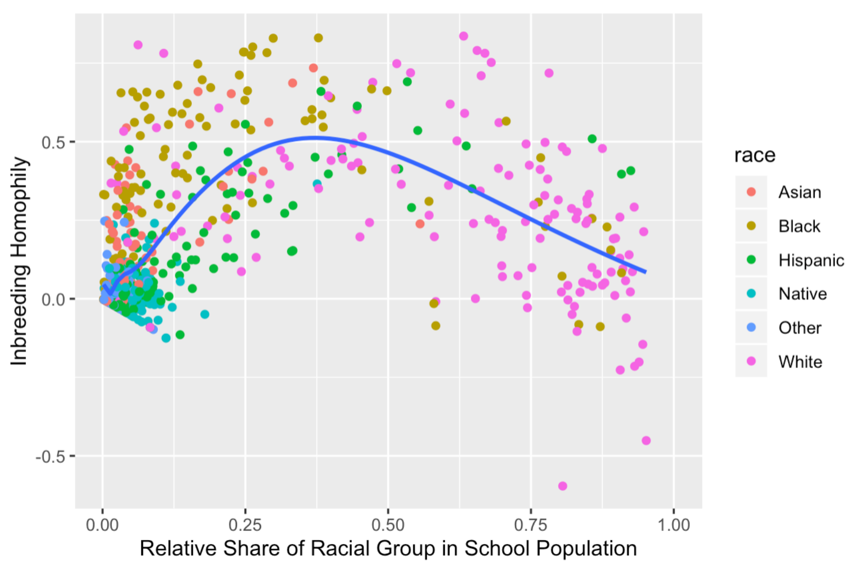
Homophily$_i$	Coefficient	Std. Error
Popshare$_i$	0.76**	0.00
constant	0.27**	0.00

** $p < 0.001$

Figure 1.5 shows the patterns of homophily for each observation, which is a racial group of a school. The vertical axis is the relative homophily index of a group i in its school, and the horizontal axis corresponds to the share of that particular group in the overall population in its school. It indicates that for most racial groups, the measure of homophily is greater than the corresponding population fraction. It means that most high school students form more friendships with their own racial group. Relative homophily is an omnipresent phenomenon in high-school friendship networks in Add Health data.[1] It can also be seen that the majority groups have higher levels of homophily, as shown by the positive and statistically significant coefficient from the regression of homophily measure on population shares in Table 1.1.

As expected, homophily increases with the share of the racial group in the school population. To see the excess level of homophily of racial groups, we also plot the Inbreeding Homophily measure of groups. The measure of Inbreeding Homophily is plotted in Figure 1.6. There is a significant and distinct pattern of inbreeding homophily as a function of relative group size. Table 1.2 presents the regression output of inbreeding homophily on a quadratic form of population sizes. As can be seen, the coefficient of the squared term is negative and statistically significant.

Figure 1.6: Patterns of Inbreeding Homophily by Population Share of Racial Groups



[1] Echenique et al. (2006) develop another segregation measure- called Spectral Index - and test it with the same data. Their figure also implies very similar results.

Table 1.2: Output for Regression of Inbreeding Homophily on Population Shares

InbreedHomophily$_i$	Coefficient	Std. Error
Popshare$_i$	2.19**	0.00
Popshare$_i^2$	−2.37**	0.00
constant	0.04**	0.00

*** $p < 0.001$

Two clear patterns can be seen in Figure 1.6. Firstly, there is inbreeding homophily for most of the racial groups. Furthermore, there is a nonlinear and non-monotone relationship as the inbreeding homophily is at its maximum in the middle-sized race groups. In contrast, it is distinctively smaller for extremely large or small ones.

1.4.3 Costly Behavior

Here we demonstrate the relation between high-school students' racial group shares in their school population and their costly behavior adoption choices. We use the following regression model for estimation:

$$CostlyBehavior_i = \beta PopulationShare_i + \sum_j \alpha^j X_i^j + \epsilon_i,$$

where we regress the costly behavior choice of student i on the population share of i's racial group in her school, and on other controlling factors that may have affected the costly behavior choice of student i.

In the model, $CostlyBehavior_i$ is defined as the following: it takes the value of $smoking_i = 1$ if student i reported smoking on one or more days in the past 30 days, and 0 otherwise. Hence, the students who reported smoking on one or more days in the past 30 days were considered current smokers, whereas all other students were considered non-current smokers. The latter group also includes people who never smoked and past smokers.[2] As robustness checks, we alternatively define costly behavior more precisely: student i is said to be a smoker if she smokes every week or daily. We follow the same procedure for $drinking_i$, $drunk_i$ and $racing_i$.

$PopulationShare_i$ is the relative share of the racial group of student i in her school population. We alternatively use $Minority_i$, a dummy variable taking 1 if student i belongs to a minority race group in her school and 0 otherwise.

X^j's are the set of controls that we include as other factors that may affect the costly behavior choice of students. Here we have $school_i$ as school identification to control for school-fixed effects. We also include $race_i$, a categorical variable that shows a student's ethnicity, age_i and $gender_i$, which is 1 for male students and 0 for females. As a proxy for economic status, we include $parenteduc_i$, the highest educational attainment by either parent.

We expect that the population share of the student's racial group has a negative effect on the costly activity of that student, as our central hypothesis is that students from small groups use these costly activities to socialize with their same kind friends. So we expect the coefficient β to be statistically significant and negative.

Estimation

We estimate using a Linear Probability Model and Probit regression since the dependent variable can take only two values: being a smoker or non-smoker.

The detailed tables of regression outputs for the dependent variables smoking, drinking, getting drunk, and racing are included in Appendix A.7. We summarize the essential points in Table 1.3.

[2] The way of defining smokers as people who smoked in the last 30 days is a conventional definition of smoking or substance use in public health literature (Alexander et al., 2001; Steuber and Danner, 2006).

Table 1.3: Regression Output of Population Share of High-School Racial Groups with Students' Costly Behavior

	Dependent Variable: Smoking				Dependent Variable: Drinking			
	Linear Prob. Model		Probit		Linear Prob. Model		Probit	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Population Share	−0.04***	0.01	−0.19***	0.04	−0.02*	0.01	−0.09**	0.04
Minority Dummy	0.007	0.005	0.05**	0.02	0.002	0.7	0.01	0.02

	Dependent Variable: Getting Drunk				Dependent Variable: Racing			
	Linear Prob. Model		Probit		Linear Prob. Model		Probit	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Population Share	−0.02**	0.003	−0.14***	0.05	−0.04***	0.00	−0.12***	0.04
Minority Dummy	0.00	0.00	0.03	0.02	0.017***	0.00	0.06***	0.02

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Our findings indicate that relative shares of students' racial groups within schools are essential in understanding adolescent costly behavior patterns. Specifically, belonging to a smaller racial group increases the likelihood of the student adopting a costly behavior. More precisely, if we take two black students from two different schools, the one from a school where black students are a minority is more likely to smoke than the one from a school where black students are a majority.

The regressions in which we use stricter definitions of the costly behavior also give a statistically significant and negative coefficient of population share, so the negative effect of population share of a racial group on the costly behavior of the students is permanent. The regression outputs are included in Appendix A.8.

1.5 Conclusion

In this paper, we propose a friendship formation model that can explain the social patterns observed in Figure 1.6 in which most groups, no matter what size, have inbreeding homophily. For that, we consider a continuum population of individuals composed of two different communities that can be categorized according to some exogenous factor such as race. Individuals decide how to allocate their time to two different social activities in which they get to meet with people of their own group and the other group.

Our finding is that there is an equilibrium in which agents from the minority group can differentiate themselves from others through the rather costly social activity to meet with their own kind and have homophilous friendships as a result. This equilibrium is shown to be unique under several specifications of the cost structure. We also empirically analyze friendship networks and costly behavior patterns of high-school students from Add Health and show that students belonging to minority racial groups in their school adopt costly behaviors more, controlling for other factors.

Our main contribution is to show that the mechanism in our setting – including social activities in the friendship formation model – induces endogenous and asymmetric socialization behaviors of a particular type, resulting in an equilibrium that matches the empirical socialization patterns. Here, our result does not mean that socialization always leads to the segregation of two types such that the minority group exhibits the costly behavior, and the larger group does not. Instead, we argue that these patterns can emerge in some circumstances due to society. Indeed, the empirical results from Add Health data show a variety of real-world compositions, and therefore it is reasonable that the theoretical model exhibits a multiplicity of equilibria.

Another contribution of this paper is to show that one aspect of high-school students' smoking/drinking problem is that teenagers use this costly activity to meet with others they prefer to be friends with. This result emphasizes the possibility of using extracurriculars as a strategically substitutable activity for smoking or drinking among high school students. Providing them with the option of several different activities can be seen as a policy tool to tackle these harmful behavior patterns of adolescents.

Bibliography

Alexander, C., M. Piazza, D. Mekos, and T. Valente (2001). Peers, schools, and adolescent cigarette smoking. Journal of Adolescent Health 29(1), 22–30.

Beyth-Marom, R., L. Austin, B. Fischhoff, C. Palmgren, and M. Jacobs-Quadrel (1993). Perceived consequences of risky behaviors: Adults and adolescents. Developmental Psychology 29(3), 549.

Bramoullé, Y., S. Currarini, M. O. Jackson, P. Pin, and B. W. Rogers (2012). Homophily and long-run integration in social networks. Journal of Economic Theory 147(5), 1754–1786.

Bramoullé, Y. and B. W. Rogers (2009). Diversity and popularity in social networks.

Clark, A. E. and Y. Lohéac (2007). "it wasn't me, it was them!" social influence in risky behavior by adolescents. Journal of Health Economics 26(4), 763–784.

Currarini, S., M. O. Jackson, and P. Pin (2009). An economic model of friendship: Homophily, minorities, and segregation. Econometrica 77(4), 1003–1045.

Currarini, S., J. Matheson, and F. Vega-Redondo (2016). A simple model of homophily in social networks. European Economic Review 90, 18–39.

Echenique, F., R. G. Fryer Jr, and A. Kaufman (2006). Is school segregation good or bad? American Economic Review 96(2), 265–269.

Evans, W. N., W. E. Oates, and R. M. Schwab (1992). Measuring peer group effects: A study of teenage behavior. Journal of Political Economy 100(5), 966–991.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. Annual Review of Sociology 27(1), 415–444.

Millstein, S. G., A. C. Petersen, and E. O. Nightingale (1994). Promoting the health of adolescents: New directions for the twenty-first century. Oxford University Press, USA.

Steuber, T. L. and F. Danner (2006). Adolescent smoking and depression: which comes first? Addictive behaviors 31(1), 133–136.

Chapter 2

Impact of Information Dissemination Strategies on Vaccine Decision: Empirical Evidence from Turkey

Abstract

In this paper, we empirically analyze a unique data set obtained from public family health centers in the city of Izmir in Turkey, focusing on attitudes and information sources regarding COVID-19 vaccines. By leveraging various communication strategies employed by family care physicians —from distributing informational pamphlets to one-to-one consultations and phone calls— we examine their effectiveness in disseminating vaccine-related information. In particular, the study categorizes these communication strategies into two primary classes: "broadcast" and "personalized." We find that "personalized" information dissemination strategies, such as face-to-face consultations and phone calls, are more effective in educating individuals about COVID-19 infection and vaccines. However, more information does not directly translate into higher vaccination rates. Nevertheless, some types of personalized modes are shown to be effective in generating desired actions only within some subgroups of patients. Specifically, for patients who trust social sources of information, one-to-one in-person meetings with a nurse demonstrate a statistically significant impact on driving vaccination, compared to broadcasting strategies.

2.1 Introduction

The COVID-19 pandemic has altered normalcy in almost every facet of human life, creating an urgent need for effective public health strategies. Central to this struggle is the challenge of vaccine dissemination and acceptance. While vaccines serve as a cornerstone for pandemic control, vaccine hesitancy remains a significant barrier. This research aims to elucidate the role of public health officials' information dissemination strategies in shaping knowledge and consequent vaccination decisions.

Public health campaigns have historically relied on many information channels, ranging from mass media campaigns to personalized physician consultations. Yet, the efficacy of these diverging strategies in a global pandemic remains a pivotal yet under-explored area of research. We specifically address two channels of impact: the influence of personalized vs. broadcast strategies on the patient's level of information, and their impact on the vaccination decision.

To conduct this empirical analysis, we leverage a unique data set from family health centers in Izmir, Turkey. The data offers insights into various attitudes and sources of information concerning COVID-19 vaccines. We differentiate between communication strategies such as distributing pamphlets, one-to-one meetings, and phone calls, categorizing them between the classes of "broadcasting" and "word-of-mouth/seeding," to determine the most effective.

Our research employs rigorous statistical methods to analyze the data set, aiming to isolate the influence of dissemination strategies from confounding variables. This approach allows us to identify the most effective strategies and explore other non-informational aspects that may affect vaccination decisions.

The study builds upon the existing body of literature in behavioral economics, public health, and information theory. Specifically, Liester (2021) reviews several methods that public health professionals consider to combat low vaccination rates.

These methods vary from education and persuasion to stricter types such as incentivization and coercion. Our research intersects with the strands of persuasion and education designed to mitigate COVID-19 vaccine hesitancy by exploring the impacts of different information dissemination strategies on vaccination decisions.

The role of communication strategies in influencing health decisions has garnered increasing academic attention. An early yet burgeoning literature focuses explicitly on the strategies to overcome vaccine hesitancy. There is no clear consensus in the literature on which public health intervention methods are effective. For instance, Bahety et al. (2021) delve into the effectiveness of text-based interventions in COVID-19 preventive behaviors in India, and finds that an SMS-based information campaign broadcasted to the general population has no evident positive impact on knowledge or adoption of preventive health behavior. On the other hand, the empirical evidence suggesting that face-to-face information or education may improve vaccination status or knowledge, is weak and non-conclusive, as shown by the meta-analysis made in Kaufman et al. (2018). More generally, Bavel et al. (2020) synthesize social and behavioral science findings from prior literature on topics relevant to pandemics, including work on social and cultural influences on behavior, and science communication.

The mechanisms through which information flows in social and professional networks have significant implications on the behaviors and decisions of individuals. Seminal work by Banerjee et al. (2018) offers experimental evidence highlighting the nuances of information dissemination during India's demonetization. Making a distinction between seeding and broadcasting the news about the demonetization's official rules widely, they show that the performance of the two types of communication depends on whether the identity of the initially informed was publicly disclosed or not. Within this line, a large-scale messaging campaign is shown to increase the health-preserving behavior of the people (Banerjee et al., 2020) and another large-scale set of interventions to increase demand for immunization shows that SMS reminders combined with incentivization perform the best among other alternatives (Banerjee et al., 2021). Furthermore, research by Breza and Chandrasekhar (2019) investigates the role of social networks in fostering reputational and commitment effects, contributing to our understanding of how these networks may potentially influence health behaviors. The current study situates itself within these discussions by examining the effectiveness of various dissemination strategies in the specific context of vaccine information.

This paper concentrates on an empirical analysis using a unique data set obtained from public family health centers in the city of Izmir in Turkey, focusing on attitudes and information sources regarding COVID-19 vaccines. By leveraging various communication strategies employed by family care physicians —from distributing informational pamphlets to one-to-one consultations and phone calls— we examine their effectiveness in disseminating vaccine-related information. In particular, the study categorizes these communication strategies into two primary classes: "broadcasting" and "personalized." This categorization serves as the analytical framework for evaluating which method most effectively influences both the level of information received by patients and their subsequent vaccination decisions.

The results show that different communication strategies impact the level of information and decisions related to the COVID-19 vaccine. We find that "personalized" modes of information dissemination, such as face-to-face consultations and phone calls, are more effective in educating individuals about COVID-19 disease and vaccines. However, more information does not directly translate into desired actions, such as higher vaccination rates. We show that there is a general lack of empirical evidence on personalized information dissemination strategies being able to nudge people into the right action or to *move the needle*. Interestingly, *some* types of personalized modes are shown to be effective in generating desired actions only within some subgroups of patients. Specifically, for patients who trust social (or *non-professional*) sources of information, personalized one-to-one meetings with a nurse demonstrate a more substantial influence on driving people to vaccinate.

The organization of the paper is as follows. The data is introduced and explained in detail in Section 2.2, and then the empirical model and estimation methods are presented in Section 2.3. Afterward, in Section 2.4, we show the findings of the empirical analysis and discuss the results. Finally, we conclude in Section 2.5.

2.2 Data

Here we give information on the data set we used in our analysis, mentioning the data collection methodology and sample selection. Also, we explain how we construct our metric of vaccine hesitancy which is used in our analysis below.

2.2.1 Research Design and Setting

This study uses a cross-sectional analytical framework to explore the vaccine hesitancy landscape in Izmir. The study was conducted in seven selected public health centers within the healthcare domain. These units are a representative sub-sample of the 1102 family medicine centers permeating the city of Izmir.

2.2.2 Population and Sample Selection

The encompassing population for this study was drawn from the patients attending the 1102 family medicine centers in Izmir. The sample, however, was more refined. Individuals who seek medical services from the pre-selected seven family medicine units during the earmarked time frame which is between January and February 2022, form the sample.

Eligibility for participation hinged on several criteria. First, individuals aged between 20 and 45 years were given primacy. Their engagement with the designated family medicine units during the study window and willingness to participate in the research were also crucial. On the other hand, individuals facing literacy challenges, those with communication barriers, or those diagnosed with psychotic disorders or dementia-related conditions were systematically excluded from the study.

With precision and rigor, the sample size was calibrated to 387. This number was predicated on achieving a confidence interval of 95%, with an underlying 50% prevalence and a 5% margin of error. The sampling strategy, rooted in convenience sampling, was purposefully designed to encapsulate a broad spectrum of willing participants. The research eventually progressed with a total of 197 participants, each meticulously satisfying the criteria mentioned.

2.2.3 Data Collection Methodology and Instruments

The data collection process focuses on a comprehensive face-to-face interview, mediated through a structured questionnaire. The questionnaire consists of two main parts aiming for clarity and depth of information. The translated version of the detailed survey form is included in Appendix B.1.

The initial part is demographic-centric, obtaining information about participants' age, gender, marital status, educational background, occupations, health center affiliations, chronic disease, and medication dependencies.

The latter section adopts a more thematic approach, aimed at exploring the multifaceted realm of COVID-19. It probes into participants' past encounters with the virus, vaccination records, and motivations, experiences within their close circles, habitual healthcare behaviors, and their informational ecosystem regarding the virus and vaccination. This section is particularly exhaustive, drawing from qualitative and quantitative strands of inquiry, all designed based on the prevailing literature.

The descriptive properties of the data are included in Appendix B.2.

2.2.4 Vaccine Hesitancy

In our data, we used the definition of vaccine hesitancy provided by the World Health Organization (WHO) and widely accepted in the medicine literature, which describes it as "the delay in acceptance or refusal of vaccination despite the availability of vaccination services," SAGE Working Group (2014). To determine the baseline for vaccine compliance, we consulted official records to pinpoint the timeline of vaccine availability for different age groups.[1] From our investigation, it was clear that by the survey date, all survey respondents (aged between 20-45) should have received three doses of the COVID-19 vaccine. We thus developed a categorical variable to capture this phenomenon as follows:

- If the respondent had taken three or more doses, she was assigned a value of 2, which indicates adherence to the recommended vaccine schedule.
- If she had not taken the minimum number of doses but had received some positive number of doses, she received a value of 1, indicating hesitancy or delay.
- If she had not taken any dose, she received a value of 0, which indicates rejection.

[1] Public records of vaccine availability in Turkey by date and age groups are available at https://turcovid19.com/etkinlikler/turkiye-covid19-asilama-gruplari-asi-uygulama-tarihleri/

2.3 Estimation

This part explains the empirical model of the paper, focusing on the two model equations. It also mentions several methods used for estimating the model.

2.3.1 Regression Equation

To see the impact of various information dissemination strategies adopted most frequently by doctors on the level of information received by the patient, the following regression equation is used:

$$\text{LevelofInformed}_i = \beta \text{InfoStrategy}_i + \alpha \text{DoctorID}_i + \sum_j \alpha^j X_i^j + \epsilon_i \tag{2.1}$$

Similarly, for the impact of information dissemination strategies on the resulting vaccine decision of the patient, the following equation is used:

$$\text{VaccineHesitancy}_i = \beta \text{InfoStrategy}_i + \alpha \text{DoctorID}_i + \sum_j \alpha^j X_i^j + \epsilon_i \tag{2.2}$$

where control variables X_i^j's include age, gender, employment status, chronic illness, the number of consultations with the doctor, if she ever had COVID-19 if anyone close to her ever died due to COVID-19. The $DoctorID_i$ is an anonymized identification of the family physician who is assigned to the patient by the public health system.

Our primary focus is on the variable $InfoStrategy_i$ which defines the most frequent communication method between the patient and the family physician about the COVID-19 disease and vaccine. It is a self-reported measure by the patient, chosen among the options: *face-to-face*, *phone*, *with nurse*, *information pamphlet*, *SMS* and *others*. Following the focus of this paper, different information strategies are grouped into two strategy classes: *Personalized* and *Broadcasting* strategies. Whereas the options of *face-to-face*, *phone* and *nurse* are regarded as *Personalized* because of their interactional, one-to-one and personalized nature, *information pamphlet*, *SMS* and *others* are grouped in *Broadcasting* strategies as in these, everyone received standardized and identical information about the COVID-19 pandemic.

The $LevelofInformed_i$ is a 0-9 scaled measure reported by each patient respondent on the level of COVID-19-related information supplied by the family physician. We consider three different topics of such information and deploy each one of them in separate regressions. These topics of information are the following: severity and course of COVID-19 disease, information about the COVID-19 vaccine, and where to get COVID-19 vaccine.

For the $VaccineHesitancy_i$ variable, we use several different measures to capture the vaccine hesitancy level of the patients in Equation 2.2. The initial measure is the number of COVID-19 vaccinations received by the patient. This number is disclosed by the patients themselves in the survey form. Alternatively, we also define a binary measure of Vaccine Hesitancy indicating if the patient is vaccine-hesitant (0) or vaccinated (1). The detailed construction method of this measure is explained in Section 2.2.4 above.

2.3.2 Estimation Methods

Different methods of estimation have been applied to the regression equations below. For Equation 2.1, we use Ordinal Logistic Regression, since the level of information is regarded as an ordered nominal variable, varying between levels 0 and 9.

For Equation 2.2, we use Poisson regression when we use the number of vaccines as the dependent variable. Additionally, when the vaccine hesitancy measure is used as the dependent variable specified in Equation 2.2, a Probit regression is used as an estimation method.

2.4 Results and Discussion

In this part, we show the results of our empirical analysis. Firstly, we report the regression results regarding the effect of information dissemination strategies on the patients' information level about the COVID-19 pandemic and the vaccination decision. Lastly, we mention how some of the results change when we put our focus on specific characteristics of people, for instance, trust in unprofessional social sources of information.

2.4.1 Information Strategy and Level of Information

Here, we present our findings from the regression of information dissemination strategies on the patients' information level, explained by the Equation 2.1. We summarize the results in Table 2.1, where the dependent variables are information level about COVID-19 disease *and* vaccine. For the sake of brevity, we report the findings of the dependent variable *where to get COVID-19 vaccine* in Table B.6 in Appendix B.3.[2]

Moreover, for each dependent variable, we present four different regression model specifications to check the robustness of our point estimates. Each specification varies across the control variables that are included in the estimation.

Looking at Table 2.1, the most prominent observation is the following: across all different regression model specifications, personalized information dissemination strategies are better than broadcasting ones in generating higher levels of information in patients. This is evident by the positive and statistically-signification

[2]The findings in Table B.6 are in line with the main findings we report in this part.

Table 2.1: Results on Information Level

Dependent variable:	*Info. Level about COVID-19 disease*				*Info. Level about COVID-19 vaccine*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Info strategy class (ref: Broadcasting)								
Personalized	2.981***	2.929***	2.526***	2.585***	3.188***	3.126***	3.074***	2.992***
	(0.588)	(0.593)	(0.806)	(0.820)	(0.598)	(0.602)	(0.842)	(0.854)
Doctor ID	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age (ref: age 20-25 years)								
25-30 years		0.069		-0.525		0.193		-0.248
		(0.764)		(0.873)		(0.789)		(0.899)
30-35 years		0.238		-0.234		0.727		0.607
		(0.775)		(0.906)		(0.804)		(0.927)
35-40 years		0.374		0.238		0.516		0.525
		(0.773)		(0.880)		(0.797)		(0.902)
41-45 years		-0.509		0.034		-0.118		0.565
		(1.170)		(1.345)		(1.171)		(1.353)
Male		-0.034		0.196		0.181		0.534
		(0.344)		(0.427)		(0.342)		(0.427)
Unemployed		-0.263		-0.379		0.018		0.048
		(0.351)		(0.427)		(0.347)		(0.425)
Chronic illness		0.695*		0.501		0.770**		0.448
		(0.391)		(0.490)		(0.391)		(0.494)
Had COVID-19			0.137	0.085			0.323	0.307
			(0.375)	(0.387)			(0.371)	(0.386)
Close death from COVID-19			0.712**	0.054			0.590*	0.326
			(0.358)	(0.486)			(0.344)	(0.477)
Numb. of consultations			0.358***	0.349***			0.264***	0.279***
			(0.123)	(0.119)			(0.098)	(0.098)
Observations	173	168	131	127	173	168	131	127

Note: *p<0.1; **p<0.05; ***p<0.01

coefficients of the personalized strategy class, compared to the reference level of the broadcasting strategy class.

The interpretation of the coefficients is not entirely straightforward since we employ an ordered logistic regression. That is why the odds ratios are calculated in Table 2.2.

Table 2.2: Odd Ratios - Results on Information Level

	(1)	(2)	(3)	(4)
Dep. variable: Info. Level about COVID-19 disease				
Info strategy class (ref: Broadcasting)				
Personalized	19.71	18.71	12.50	13.26
Dep. variable: Info. Level about COVID-19 vaccine				
Info strategy class (ref: Broadcasting)				
Personalized	24.24	22.78	21.63	19.93

Let us look at specification (1). As the reference level of strategy classes is *broadcasting*, the interpretation of the coefficient is as follows: For patients who received information in *personalized* ways, the odds of having more information about COVID-19 disease is 19.7 times that of patients who received broadcasting information. Similarly, their odds of having more information about COVID-19 vaccine is 24.2 times that of others.

It is an intuitive finding as the personalized information dissemination strategies employed by the doctors may have an external positive impact on building trust and making the conversation relevant to the patients, even though the content of the message is probably the same across different strategy types. This finding aligns with the importance of interpersonal communication in healthcare on increasing health-related information in patients (Chichirez and Purcărea, 2018; Berry et al., 2003; Kreuter and Wray, 2003). As a result of this positive impact, patients form a subjective belief that they are more informed about some aspects of the conversation, in this case about COVID-19 disease and vaccination.

Another insight from the results in Table 2.1 is that people with chronic illnesses feel more informed by their physicians regarding COVID-19, evident by the statistically significant positive estimate in the specification (2) of both of the dependent variables. This may be because people with chronic illnesses are regarded in *high-risk* group of the population regarding the COVID-19 pandemic, which is a widely-documented fact in medicine literature. (Williamson et al., 2020; Rosenthal et al., 2020; Hacker et al., 2021; Laires et al., 2021). This may lead their family physicians - or patients- to provide or seek more detailed information on this topic.

When we include the *number of consultations* as an exploratory variable to the equation - done in the specification (4) - we observe the statistically significant positive impact of the sheer quantity of interaction on the resulted information level of the recipient. For instance, for the patients who had one *more* consultation with their doctors, their odds of having more information about the COVID-19 vaccine is 1.32 times more than others, keeping everything else constant. This is intuitive since more consultations with their family physicians mean more opportunities to discuss COVID-19-related health matters. Moreover, we see that the statistically significant coefficient of chronic illness in the specification (2) becomes more minor and insignificant in specification (4) after including the number of consultations in the equation. It can be because people with chronic diseases visit the doctor more often. When we add the term for the sheer number of visits, it clears out the effect of frequent visits from chronic illnesses and reduces its coefficient, even though the resulting coefficient remains positive.

A further observation from the results is the impact of the personal life experiences of the patient on their information level on COVID-19. For instance, the coefficient of the binary variable capturing the information on whether anyone close to the patient had died because of COVID-19 appears positive and statistically significant. Particularly, holding everything else constant, the people who experienced the death of a person that is familiar to them, have odds of having more information about COVID-19 disease 2.04 times more than that of others who had not experienced a close person's death. This result makes sense as someone who has gone through such a loss due to COVID-19 may seek more information about this disease and receive more information from their family physician.

As a last remark, there is no clear and statistically significant pattern on the impact of patient's age and gender differences on their information level of COVID-19 disease and vaccine. However, the coefficients of older age groups appear primarily positive. This may signal an increased information level for older patients. Still, because the coefficients all appear non-significant, we can conclude that age has no apparent impact on this matter.

2.4.2 Information Strategy and Vaccine Decision

In the former part, we show that personalized information dissemination strategies of family physicians positively impact people's information level regarding COVID-19 matters such as the disease itself and vaccination. However, a further question can be the following: Do the personalized strategies impact people's decision to have the COVID-19 vaccine or not? Providing people with as much correct information as possible is one thing, but the ultimate aim of health professionals is to steer people into the correct health behavior choice, in this case, to vaccinate.

With this aim, in this part, we show the findings from the regression of information dissemination strategies on the patients' vaccination decisions, formulated in Equation 2.2. We summarize the results in Table 2.3, where the dependent variables are the vaccine hesitancy measure and the total number of COVID-19 vaccines. Similar to the part above, for each dependent variable, we have four different specifications of the model where they differ on the control variables included.

Focusing on Table 2.3, the first observation is that personalized information dissemination strategies have statistically insignificant and negative coefficients across all specifications of the models and two different dependent variables of vaccine decision. It shows the lack of evidence for our initial question on the effectiveness of personalized interventions of healthcare professionals. The odds ratios are reported in Table 2.4 to make interpreting the results easier.

Table 2.4: Odd Ratios - Results on Vaccine Decision

	(1)	(2)	(3)	(4)
Dep. variable: Vaccine Hesitancy				
Info strategy class (ref: Broadcasting)				
Personalized	0.62	0.66	0.54	0.80
Dep. variable: Number of COVID-19 vaccine				
Info strategy class (ref: Broadcasting)				
Personalized	0.88	0.89	0.81	0.89

Looking at the coefficient of a binary variable showing if the person had COVID-19 before or not, having had COVID-19 (compared to not having had) decreases the odds of being *more* vaccinated. It may be because having had the disease already makes them immunized to the virus, and they are not advised to get the vaccine for some period after their initial infection.

Furthermore, the results show that being unemployed (compared to the employed reference) decreases the odds of being more vaccinated although the coefficient remains insignificant in all specifications. This is reasonable as workers are arguably more exposed to social interactions hence the virus, compared to unemployed ones. It may lead them to be fully vaccinated in more frequent cases. Also, having a close

Table 2.3: Results on Vaccine Decision

Dependent variable:	*Vaccine Hesitancy with three levels*				*Number of COVID-19 vaccine*			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Info strategy class (ref: Broadcasting)								
Personalized	-0.479	-0.409	-0.619	-0.281	-0.137	-0.115	-0.206	-0.119
	(0.521)	(0.542)	(0.704)	(0.759)	(0.165)	(0.168)	(0.229)	(0.238)
Doctor ID	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age (ref: age 20-25 years)								
25-30 years		1.897**		2.464**		0.414		0.497
		(0.886)		(1.001)		(0.303)		(0.314)
30-35 years		1.785**		1.723*		0.444		0.405
		(0.901)		(1.018)		(0.310)		(0.325)
35-40 years		2.025**		2.555**		0.419		0.495
		(0.905)		(1.020)		(0.310)		(0.318)
41-45 years		2.477**		2.662*		0.527		0.628
		(1.189)		(1.477)		(0.406)		(0.461)
Male		0.213		0.244		0.057		0.077
		(0.377)		(0.480)		(0.113)		(0.138)
Unemployed		-0.315		-0.174		-0.152		-0.076
		(0.371)		(0.461)		(0.119)		(0.137)
Chronic illness		-0.038		0.062		-0.023		-0.043
		(0.412)		(0.511)		(0.135)		(0.162)
Had COVID-19			-0.038	0.211			-0.053	-0.024
			(0.394)	(0.425)			(0.124)	(0.127)
Close death from COVID-19			0.025	0.169			0.038	0.040
			(0.471)	(0.510)			(0.111)	(0.150)
Numb. of consultations			0.075	0.040			0.027	0.017
			(0.104)	(0.109)			(0.031)	(0.032)
constant					1.145***	0.731*	1.164***	0.657
					(0.207)	(0.389)	(0.269)	(0.444)
Observations	176	171	131	127	176	171	131	127
Log Likelihood					-280.804	-269.563	-206.112	-197.207
Akaike Inf. Crit.					579.608	571.125	436.224	432.415

Note: *p<0.1; **p<0.05; ***p<0.01

one die due to COVID-19 increases the odds of being fully vaccinated, which makes sense as it can make people more aware of the risks and complications of the virus leading them to take precautions against it such as vaccination.

In any case, we can circumspectly say that the insignificant and negative coefficients of personalized strategies imply that physicians' personalized information dissemination techniques do *not* have an evident positive impact on the consequent vaccine decision of patients. It is in line with several papers in the literature mentioning there is no firm evidence of any specific intervention's ability to affect vaccine hesitancy/refusal (Dubé et al., 2015; Nyhan et al., 2014).

2.4.3 Information Strategy and Vaccine Decision by Most Trusted Information Source

In this part, we explore the possibility of seeing the impact of information dissemination strategies on people's vaccination decisions by focusing on a sub-sample of people who are more *prone* to non-professional sources of COVID-19-related information. It is a relevant distinction because the overall sample may mainly consist of people who regularly see their family physicians (as the questionnaire was conducted at the entrance of family medicine centers), highly trust their doctors, and follow their instructions on health behavior. On the other hand, focusing on the subgroup of the population who claim to have trust in other non-medical social sources may give a more accurate idea of the effectiveness of different medical information dissemination strategies as these people will be the relevant group of focus in a possible health-related public intervention.

Based on this argument, we re-run our estimation shown in Equation 2.2, but now focusing only on the sample of people who trust social (and non-professional) sources of information on the COVID-19 pandemic. This is done by using one question in the survey, asking people to pick out their most trusted three sources of information about the COVID-19 vaccine.

As can be seen in Table B.4, most of the responses are professional sources of information such as *scientists* or *doctors*. The people who reported at least one of the social sources of information (more specifically, the options *internet, social media, TV/newspaper, family, friends, and neighbors*) as their most trusted source are grouped in a sub-sample. The regression output is reported in Table 2.5. [3]

The main interesting observation is the following: across all different specifications of the regression model, each personalized information dissemination strategy has different impacts on people's vaccination decisions, compared to the broadcasting strategies. Specifically, we see that *meeting with a nurse* and *phone-call* are better than broadcasting strategies in generating higher levels of vaccination, across all model specifications. Moreover the estimate of *meeting with a nurse* remains statistically significant.

Table 2.6: Odd Ratios - Most trusted social media

	(1)	(2)	(3)	(4)
Dep. variable: Vaccine Hesitancy				
Info strategy class (ref: Broadcasting)				
Face-to-face	0.50	0.24	0.55	0.77
Nurse	3.54	4.42	3.63	3.09
Phone	1.35	1.57	1.45	1.15

Looking at the odds ratios in Table 2.6, in the specification (1), the results show that for patients who received information by meeting with a nurse, the odds of being more vaccinated against COVID-19 disease are 3.54 times that of patients who

[3] Similar estimation is done for the subgroup of people who report at least one professional source of information as their most trusted source. In this case, the estimates are almost identical to the ones done with the whole sample as almost all of the sample remained in this subgroup. Therefore, we omit reporting these regression outputs for the sake of brevity.

received broadcast information. Likewise, for the patients who received information over a phone call with the physician, the odds of being more vaccinated are 1.35 times that of others who received information via broadcasting.

Nurses often serve as the front-line health professionals in various settings, making them the primary point of contact for patients. Their accessibility, approachability, and perceived relatability often facilitate more open, candid conversations about health concerns. This ease of communication may be especially valuable in addressing vaccine hesitancy, where misunderstandings and misbeliefs can significantly influence decisions. As trusted medical professionals who are often less intimidating than physicians, nurses may have a better capacity to engage with patients in a manner that encourages dialogue rather than mere dissemination of information, especially when there is skepticism about vaccine effectiveness or lack of trust in those recommending the vaccination.

2.5 Conclusion

Our empirical study set out with dual objectives: firstly, to examine the impact of different communication strategies—namely "broadcasting" and "personalized"—in information dissemination about COVID-19 vaccines, and secondly, to understand how these strategies influence individual vaccination decisions. The research leveraged a unique data set collected from public family health centers in Izmir, Turkey, yet aimed to offer insights of broader relevance.

Our findings underscore that personalized modes of communication, such as one-to-one consultations and phone calls, are more powerful in enhancing individuals' informational levels about COVID-19 and vaccines. However, a critical nuance emerges where increased information does not necessarily correlate with an uptick in vaccination rates. While personalized strategies effectively educate individuals, they do not uniformly catalyze action toward vaccination. Interestingly, subgroups exist where personalized interactions were highly influential; notably among those who trust social or non-professional information sources, one-to-one consultations with nurses were particularly effective in promoting vaccination.

This research illuminates the complex relationship between information dissemination and vaccine hesitancy and nuances of the effectiveness of personalized communication, thereby aiming to provide a richer framework for public health strategy formulation of policy-makers.

Ultimately, vaccine hesitancy is not solely a byproduct of insufficient or ineffective communication; it is influenced by numerous determinants, ranging from cultural beliefs and past medical experiences to trust in healthcare systems. While communication strategies, whether personalized or broadcast, are essential elements, they are part of a larger ecosystem of factors that influence health behaviors.

Bibliography

Bahety, G., S. Bauhoff, D. Patel, and J. Potter (2021). Texts don't nudge: an adaptive trial to prevent the spread of covid-19 in india. Journal of Development Economics 153, 102747.

Banerjee, A., M. Alsan, E. Breza, A. G. Chandrasekhar, A. Chowdhury, E. Duflo, P. Goldsmith-Pinkham, and B. A. Olken (2020, July). Messages on covid-19 prevention in india increased symptoms reporting and adherence to preventive behaviors among 25 million recipients with similar effects on non-recipient members of their communities. Working Paper 27496, National Bureau of Economic Research.

Banerjee, A., E. Breza, A. G. Chandrasekhar, and B. Golub (2018). When less is more: experimental evidence on information delivery during india's demonetization. Technical report, National Bureau of Economic Research.

Banerjee, A., A. G. Chandrasekhar, S. Dalpath, E. Duflo, J. Floretta, M. O. Jackson, H. Kannan, F. N. Loza, A. Sankar, A. Schrimpf, et al. (2021). Selecting the most effective nudge: evidence from a large-scale experiment on immunization. Technical report, National Bureau of Economic Research.

Bavel, J. J. V., K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M. J. Crockett, A. J. Crum, K. M. Douglas, J. N. Druckman, et al. (2020). Using social and behavioural science to support covid-19 pandemic response. Nature Human Behaviour 4(5), 460–471.

Berry, D. C., I. C. Michas, and E. Bersellini (2003). Communicating information about medication: the benefits of making it personal. Psychology and Health 18(1), 127–139.

Breza, E. and A. G. Chandrasekhar (2019). Social networks, reputation, and commitment: evidence from a savings monitors experiment. Econometrica 87(1), 175–216.

Chichirez, C. and V. Purcărea (2018). Interpersonal communication in healthcare. Journal of Medicine and Life 11(2), 119.

Dubé, E., D. Gagnon, N. E. MacDonald, et al. (2015). Strategies intended to address vaccine hesitancy: Review of published reviews. Vaccine 33(34), 4191–4203.

Hacker, K. A., P. A. Briss, L. Richardson, J. Wright, and R. Petersen (2021). Peer reviewed: Covid-19 and chronic disease: the impact now and in the future. Preventing Chronic Disease 18.

Kaufman, J., R. Ryan, L. Walsh, D. Horey, J. Leask, P. Robinson, and S. Hill

(2018). Face-to-face interventions for informing or educating parents about early childhood vaccination. Cochrane Database of Systematic Reviews (5).

Kreuter, M. W. and R. J. Wray (2003). Tailored and targeted health communication: strategies for enhancing information relevance. American Journal of Health Behavior 27(1), S227–S232.

Laires, P. A., S. Dias, A. Gama, M. Moniz, A. R. Pedro, P. Soares, P. Aguiar, and C. Nunes (2021). The association between chronic disease and serious covid-19 outcomes and its influence on risk perception: survey study and database analysis. JMIR Public Health and Surveillance 7(1), e22794.

Liester, M. B. (2021). A review of the use of persuasion and coercion to overcome covid-19 vaccine hesitancy. Journal of Peer Scientist 4(2), e1000037.

Nyhan, B., J. Reifler, S. Richey, and G. L. Freed (2014). Effective messages in vaccine promotion: a randomized trial. Pediatrics 133(4), e835–e842.

Rosenthal, N., Z. Cao, J. Gundrum, J. Sianis, and S. Safo (2020). Risk factors associated with in-hospital mortality in a us national sample of patients with covid-19. JAMA Network Open 3(12), e2029058–e2029058.

SAGE Working Group (2014). Report of the sage working group on vaccine hesitancy. World Health Organization.

Williamson, E. J., A. J. Walker, K. Bhaskaran, S. Bacon, C. Bates, C. E. Morton, H. J. Curtis, A. Mehrkar, D. Evans, P. Inglesby, et al. (2020). Factors associated with covid-19-related death using opensafely. Nature 584(7821), 430–436.

Table 2.5: Results - Most trusted social media

	Dependent variable: Vaccine Hesitancy			
	(1)	(2)	(3)	(4)
Info strategy (ref: broadcasting)				
Face-to-face	-0.699	-1.428	-0.601	-0.266
	(0.820)	(1.001)	(0.888)	(1.047)
Nurse	1.263***	1.486***	1.290***	1.128**
	(0.250)	(0.379)	(0.297)	(0.401)
Phone	0.300	0.451	0.369	0.140
	(0.919)	(1.051)	(0.942)	(1.243)
Doctor ID	Yes	Yes	Yes	Yes
Age (ref: age 20-25 years)				
25-30 years		1.682		
		(1.513)		
30-35 years		1.991		
		(1.559)		
35-40 years		2.327		
		(1.581)		
40-45 years		-1.034		
		(2.943)		
Male		-1.218*		
		(0.689)		
Unemployed		-0.030		
		(0.699)		
Chronic illness		0.216		
		(0.743)		
Had COVID-19				0.480
				(0.696)
Close death from COVID-19			0.214	0.740
			(0.605)	(0.801)
Numb. of consultation				0.009
				(0.143)
Observations	71	70	70	58

Note: *p<0.1; **p<0.05; ***p<0.01

Chapter 3

Social Learning and Degree of Coarseness in Communication

Abstract

In this paper, we explore the impact of the quality of communication—referred to as its 'coarseness'—on social learning within connected groups. Building on the existing literature, we scrutinize the conditions under which agents in a highly connected community become stuck in incorrect beliefs. We find that the limitations of this 'stuckness' are closely tied to the model's assumptions about the simplicity of communication and the binary nature of states. Additionally, we show the equivalence of a clan and 1/2-cohesiveness. We also state the true generalization of the 'stuckness' condition to the case with a discrete number of possible actions (more than two), and show how it breaks down in the limit.

3.1 Introduction

Social ties are major channels of transmitting information, behavior, and opinions. They convey information by observing other individuals' decisions together with the conversations. The information flow through social networks has a major role in various types of phenomena, such as product choice (Trusov et al., 2009), job search (Montgomery, 1991), financial planning (Duflo and Saez, 2003), voting (Beck et al., 2002) and criminal activity (Ballester et al., 2006). Because of this, it is crucial to understand how the information flows through social ties, how private beliefs and behaviors evolve over time, how this depends on the network structure, whose opinions are particularly influential, and whether or not society can aggregate the dispersed information efficiently.

In social learning literature, the earliest attempts have been made to explain the extensive conformity within a group, i.e., herding behavior. Both Banerjee (1992) and Bikhchandani et al. (1992) construct models such that agents take actions sequentially, referred to as sequential social learning models. Agents observe the history of actions and a personal (private) signal before choosing their actions. The previous actions of others give information about which action is the correct one to the agent, even though the payoff of an agent is independent of the actions made before. This fact stands as a positive externality of information. They analytically show that people usually prefer to ignore their private information and rather imitate the crowd. In this instance, agents' choices stop presenting novel information to the newcomers, and the herding may happen on a wrong action. Several variations of this idea have been studied in many contexts, such as fads, fashions, and stock price bubbles. (see an overview in Chamley (2004)). Due to the restrictions of sequential social learning models, many fundamental issues about the dynamics of individual opinions and choices in a context where agents make many decisions and repetitively influence each other cannot be studied. These restrictions force a departure from this framework and deem it necessary to study this subject in a network setting.

In the network framework, two approaches exist that differ in how they presume how people process the information they hear from each other. Assuming their particular specification on how people process the information that they receive in each period of time, with both approaches, the literature focused on answering crucial questions such as whether society as a whole converges to a unique belief in the long run, (if so) whether this is the true belief, and how fast should we expect this convergence.

One of these approaches which models the dynamics of repeated updating of beliefs assumes that people act myopic in each stage. Given the prohibitively complex form that social networks are often shaped in, it can be quite challenging for the agents involved to update beliefs properly, considering the repeated transfers of information among large numbers of individuals in many stages. This is the main assumption of the DeGroot Model developed by DeGroot (1974). DeMarzo et al. (2003) is the first paper that discussed it in detail in a microeconomic context. In their setting, agents update their beliefs or attitudes in each period simply by taking weighted averages of their neighbors' opinions from the previous period, possibly placing some weight on their own previous beliefs. The agents in their scenario are boundedly rational, failing to adjust correctly for repetitions and dependencies in the information they hear multiple times over the communications. This characterization allows us to use Markov chains' properties to find the system's long-run behavior. As long as some sufficiency conditions on the network hold, there exists a long-run consensus (Meyer and Stewart, 2023) in which the influence of each agent's initial signal on consensus can be expressed in terms of their centralities (Golub and Jackson, 2010).

Formally, Golub and Jackson (2010) derive conditions on the listening structure such that convergence to the truth - *wisdom of the crowds* - occurs. They show that all opinions in a large society converge to the truth if and only if the influence of the most influential agent vanishes as the society grows. They also identify obstructions to asymptotic learning, including excessively influential groups interrupting the long-run efficiency.

The second approach to modeling how people update beliefs is the Bayesian approach (Gale and Kariv, 2003). Agents are assumed to be fully rational, they are aware of the network and they update information via Bayes' rule. A central intuition is that each agent's belief allows her neighbors to infer what that player could have seen last period, narrowing down the set of possible states. When this process stops, it is common knowledge between any pair of neighbors what their beliefs are (Mueller-Frank, 2013, 2014). Gale and Kariv (2003) finds that if the network satisfies a connectedness assumption, the initial diversity resulting from diverse private information is eventually replaced by the uniformity of actions, though not necessarily of beliefs, in finite time with probability one. A more recent paper, Mossel et al. (2015), also studies a model of repeated interaction with fully rational expectations playing a Perfect Bayesian game, but with discounting future payoffs. They characterize the general class of networks such that players can learn the true state almost surely. Their condition of *L-local-connectedness* requires that no agent is excessively influential, with a similar intuition as its myopic-model counterparts.

Apart from the theoretical analysis, whether people in reality behave as if they are boundedly rational or fully rational is another essential question. There has been a recently growing literature on experiments on testing learning models. The laboratory experiments (Corazzini et al., 2012; Battiston and Stanca, 2014) show that people fail to consider the repetition of information even with small networks. Also, other papers show that the actual learning patterns of people are well approximated by the naive-learning approach of DeGroot (Grimm and Mengel, 2020).

On the other hand, it is only natural to consider that agents may be heterogeneous in their sophistication or naivete about how they engage in social learning. Particularly, they can vary in how well they can assess how much independent information is contained among their social connections, along with whether they account for the naivete of those connections. Mueller-Frank (2014) analyzes such a model of social networks among agents with differing degrees of sophistication and where agents can transmit their posterior beliefs to each other. He shows that at least one Bayesian agent in a strongly connected network is sufficient for perfect information aggregation.

There also exists a recent paper by Chandrasekhar et al. (2020) that considers the possible heterogeneity of agents in their sophistication of learning in a specified manner. They propose an incomplete information model of social learning with coarse communication and binary states on a network where agents can potentially be Bayesian or DeGroot. Coarse communication means that agents cannot or do not transmit their beliefs or information sets to their friends to a fine degree. They are constrained to process coarse information from their friends when they engage in learning. It can be because communication is very complex, detailed information is very costly or agents cannot do so. In this setting, coarse communication means agents observe that their friends believe the true state is 0 or 1. As a result, the paper identifies a few behavioral patterns in the learning behavior of two different types of agents and focuses on some features of the network that lead to the failure of asymptotic learning. They single out one incident where a clan of DeGroot agents remains stuck with the wrong action.

In our research, we focus on the learning properties of networks with highly clustered groups that consist of individuals with dense friendship patterns internally and sparse friendships externally. We begin by investigating how general the result from Chandrasekhar et al. (2020) about agents in a highly connected group getting stuck in the wrong state, leading to the failure of true learning in society. By using that paper as a starting point, we examine how the degree of coarseness of communication translates into social learning.

We show that the result of DeGroot agents from a clan being stuck in the wrong state is highly related to the assumptions of coarse communication and the binary state of the model. If we vary either of these assumptions slightly, agents in a clan do not get stuck in the wrong state. Indeed, the concept of a clan is equivalent to the *q-cohesiveness* which is already a well-recognized condition of clustering in social networks. We show this equivalence and we identify the condition where a clan may choose a wrong action forever when the discrete number of possible actions is more than two. Finally, it is shown that as we make communication finer by increasing the number of possible actions that people may take in each period, the case where people in a clan choose the wrong action forever becomes impossible in the limit.

The paper's organization is as follows: Section 3.2 presents the theoretical model. In Section 3.3, it is shown that the modifications we have made to show that "the stuckness of the clan" can not be generalized to different settings. Also, we show the equivalence of a clan and 1/2-cohesiveness. Then we state the true generalization of the "stuckness" condition, and show how it breaks down in the limit. In Section 3.4, we summarize the research.

3.2 Model

In this section, we describe the social learning model. We use the theoretical setting from Chandrasekhar et al. (2020) as a starting point in our results in the following section, but we generalize the coarse communication environment from binary actions to k number of possible actions. Apart from that, the structure used is as follows.

Let us consider an undirected, unweighted graph $G = (V, E)$ where V indicates the set of nodes/agents and E is the edge list of $n = |V|$ agents. $N_i = \{j : G_{ij} = 1\}$ denotes the neighborhood of i and $N_i^\star = N_i \cup \{i\}$.

It is a model of incomplete information as all agents are either DeGroot(D) or Bayesian(B) where type $\eta_i \in \{D, B\}$. Types are identically and independently distributed, and the distribution of types is standard information whereas no one knows the type of others in the network. The type describes how an agent processes information, either by using DeGroot or Bayesian updating.

Agents try to learn and correctly guess the world's underlying state $\theta \in \{0, 1\}$. Time is discrete so $t \in \mathbb{N}$. At the beginning of time, everyone receives an initial informative signal about the true state that is independently and identically distributed such that $s_i = \theta$ with some probability $p \in (1/2, 1)$.

A crucial assumption of the model is that communication is coarse. That is, in each period t, each agent i takes an action $a_i^t \in \{0, 1\}$ simultaneously which reflects his best guess about the state of the world, observes the guesses of his neighbors in the network, and updates his guess accordingly in the next period.

DeGroot agents follow the standard DeGroot updating process in a binary environment, meaning they follow most of their own guesses and their friends' guesses in the prior period. Bayesian agents try to infer what other agents could have observed last period by looking at their neighbors' choices at each period, and therefore narrow down the set of possible states. Because of the incomplete information setting, they also try to learn about the types of all other agents while learning about the state to make the most informed guess in every period.

To state their main result, mentioning the concepts of a clan and stuckness is necessary. Let d_i be the degree of agent i. A group of agents, $C \subset V$ is said to be a clan if $\forall i \in C$, $d_i(C) \geq d_i(V \setminus C)$ where $|C| \geq 2$. A clan is a set of nodes that are more connected among themselves than to those outside the group. An agent i is said to be stuck on the wrong state if there exists some time period t such that $a_i^{t+m} = 1 - \theta$ for all $m \in \mathbb{N}$. In words, i is stuck on the wrong action, if after some point in time, he always chooses that action from then on.

With these concepts defined, their central theoretical result says that if a clan consisting of all DeGroot agents agrees on the wrong state of the world at any point in time, then all the agents in the clan are forever stuck in the wrong state.

In the subsequent example, we demonstrate how a clan gets stuck in the wrong state.

3.2.1 An Example

Let there be 7 DeGroot-type agents. Let the initial signals be distributed such that Mr. 1, 3, and 6 receive a signal of 1 (indicated with a blue node) and Mr. 2, 4, 5, and 7 receive 0 (indicated with a white node). As most of the initial signals distributed are White and initial signals are assumed to be informative, the actual state of the world is White. Since the private signal is the only information they have obtained, all agents play according to their private signal in the first period. So the network in $t = 1$ is the following:

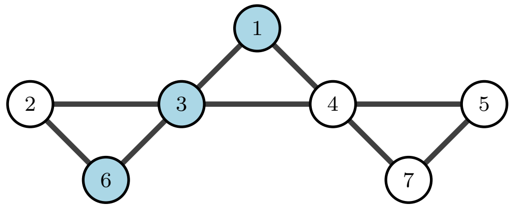


Figure 3.1: Actions at t=1

Notice that Mr. 2, 3, and 6 form a clan as all of the neighbors of Mr. 2 and 6 and half of Mr. 3's neighbors are inside this set. Let us assume that all those agents are DeGroot type. As Mr. 2 observes 2 Blue and 1 White in the first period, the majority of actions he observes are Blue and this means that he will choose the action Blue in the following period. Similarly, since Mr. 3 observes 3 Blue and 2 White, he will also continue with the action Blue in the next period. With this process followed by all agents, in $t = 2$, the choices become the following:

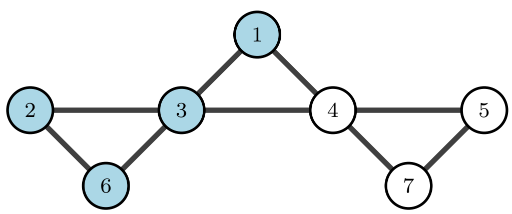


Figure 3.2: Actions at t=2

After that point, this clan remains stuck with the action Blue, seeing that most of each one's neighbors also choose Blue. Hence, they choose Blue for all $t \geq 2$ which is the wrong state.

3.3 Results

In this section, we first show our modifications to the example from the previous section and demonstrate how the clan does not get stuck in the wrong state in each case. Then, we mention the equivalence of a clan and a *q-cohesive* group and show how the "stuckness of clan" result breaks down as the number of possible actions increases. Then, we give the true condition on the "stuckness of clan" using *q-cohesiveness*.

3.3.1 Modifications on the Example

In order to show that the result of the "stuckness of clans" depends very much on the particular setting of the model - namely, coarse communication and binary state - we make three separate modifications to these assumptions of the model and demonstrate that agents in the clan do not get stuck with the wrong state anymore. Particularly, using the network configuration from the previous example, the cases are considered where

- the updating rule of agents is to choose $a_i^t = 1$ if their posterior belief that the state is 1 is greater than 2/3,
- there are more than 2 possible states/actions,
- agents can share their posterior beliefs about the states.

Firstly, we consider the case where the updating rule of agents is to choose a particular action of the two possibilities when the posterior belief about that state being the correct one is different than 1/2, for example, 2/3. It may be because one of the states is perceived to be very unlikely, so agents need a higher posterior belief about that state being the truth to take that action. In this case, it is shown that the clan can avoid getting stuck in the wrong state. The details of this case are presented in Case I of Appendix C.1.1.

Secondly, we focus on the case where the state is not binary - namely, 0 or 1 - but has more than 2 possible states. Notably, it is assumed that there are 5 different possible actions/states. We show that after all the members in the clan agree on some wrong action, they do not get stuck on this wrong action; instead, one of them can get out of that wrong state since the communication itself becomes finer with the increased number of states in the interval. This modification is shown in Case II of Appendix C.1.2.

Lastly, we concentrate on the case in which agents conduct perfectly fine communication, meaning they can transmit the posterior beliefs about the states to each other at each period. Again in this case, we show that agents can get out of the wrong state by using the results of the long-run consensus of continuous DeGroot models of Golub and Jackson (2010). The steps are explained in Case III of Appendix C.1.3.

To sum up, by considering various modifications to the model's assumptions, we show that the stuckness result depends very much on the model's binary action and coarse information assumptions. In fact, if we modify the setting slightly, the central result of the paper breaks down.

3.3.2 Equivalence of a Clan and Q-Cohesiveness

As an observation, we argue that the notion of a clan seems arbitrary. To be more specific, it is identical to *q-cohesiveness* which is a well-accepted concept of clustering made by Morris (2000). In his paper, Morris conceptualizes that a set of nodes S is *q-cohesive* concerning network G if

$$\min_{i \in S} \frac{|N_i(G) \cap S|}{|N_i(G)|} \geq q$$

Intuitively, S is q-cohesive if each node in S has at least a fraction q of their neighbors also in S. Hence, it is precisely what a clan is according to the definition of clan in the previous section, if we take $q = \frac{1}{2}$. Proposition 1 states this observation.

Proposition 3.1. *A set of agents C is a clan if and only if they are $\frac{1}{2}$-cohesive.*

Proof: See Appendix C.2.1.

3.3.3 Breakdown of "Stuckness" Result

Another observation is that the result saying that clans that consist of all DeGroot agents may get stuck in the wrong state is not related to the incomplete information setting which was argued as the novelty of Chandrasekhar et al. (2020). Instead, it is related to the coarse communication assumption.

In fact, if we were to increase the number of possible actions, but still keep it discrete, the case where clans get stuck on some given state becomes increasingly unlikely. In the extreme case, where the number of actions goes from two to infinity, the result of clans getting stuck vanishes. This is what we state in the following proposition.

Proposition 3.2. *Let k be the number of possible states/actions. As the number of possible actions/states k increases, the "stuckness of clans" result disappears in the limit.*

Proof: See Appendix C.2.2.

The detailed proof is presented in Appendix C.2.2, but the idea of the proof is as follows: for a given k, we define an interval $i^\star$ such that an agent chooses a specific action if and only if his updated belief that he forms observing his neighbors, is located at most $i^\star$ distance away from that action. After that, we show that as $k \to \infty$, $i^\star \to 0$, the incident where agents get stuck at some given action becomes impossible.

3.3.4 Generalization of "Stuckness" Condition to the Case with k Possible Actions/States

Previous results are specific to the setting where the agent can communicate only two possible states/actions ($k = 2$). In this section, we also generalize the result of agents in a clan getting stuck at the wrong state and characterize the condition for the number of possible actions $k \geq 2$.

Proposition 3.3. *Assume all agents in a $(1 - \frac{1}{2(k-1)})$-cohesive group S are DeGroot and there exist $t \geq 1$ and some fixed action a such that for all $i \in S$, $a_i^t = a$. Then $a_i^{t+\tau} = a$ for all $i \in S$ and $\tau \in \mathbb{N}$.*

Proof: See Appendix 2.3.

I should emphasize that the threshold cohesiveness level of a group depends on $i^\star$ for an arbitrary k. When $k = 2$, it corresponds to the 1/2-cohesiveness, as before. As k increases, this threshold converges to 1 very fast, which corresponds to an unrealistically high level of cohesiveness of the group. It means a group gets stuck in the wrong state if they are highly clustered, which is very restrictive and unlikely.

3.4 Conclusion

This study has delved into the complexities of social learning within highly connected groups, particularly examining how the quality of communication impacts the formation and persistence of beliefs. By extending the work of Chandrasekhar et al. (2020), we have shown that the phenomenon of agents becoming 'stuck' in incorrect beliefs is not merely a byproduct of their social environment but is intricately tied to the limitations of coarse communication and binary states.

We introduced the concept of "1/2-cohesiveness" to provide a nuanced understanding of the unity within these groups. Furthermore, we expanded the scope of the 'stuckness' condition to include scenarios with a discrete number of possible actions, revealing how this condition breaks down in the limit.

Our findings have broader implications for studying social networks, belief formation, and information dissemination. They highlight the need for more nuanced models to capture human interaction and learning complexities. Future research could focus on how different types of networks or varying degrees of communication coarseness could impact social learning. Additionally, empirical studies could be conducted to validate the theoretical models presented here.

A natural next step would be to focus on coarse communication but with more than two possible actions and attempt to find general conditions that give rise to asymptotic learning, in which the beliefs of all agents in the social network converge over time and agents take correct actions. We should emphasize that stuckness is only one of the conditions that would disrupt the asymptotic learning, there can still be other obstacles that stop people from converging to a true consensus.

For instance, let us assume that we eliminate the cases in which agents get stuck in some wrong state. Agents still need to have enough friends to observe information from a sufficient number of sources. It is crucial because it would lead them to update their beliefs smoothly enough toward the correct action (which is also the most common signal in the whole network), despite the granularity of the action set. Hence, we infer that there is a condition for asymptotic learning on the number of connections that depend on k (cardinality of action set).

Another aspect would be how convergence speed depends on the cardinality of the action set as discussed earlier. Intuitively, we expect the convergence to happen faster as the number of possible actions increases. The reason is that as the cardinality of the action set increases, its granularity decreases and it becomes easier for agents' actions to move toward the true choice.

Bibliography

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). Who's who in networks. wanted: The key player. Econometrica 74(5), 1403–1417.

Banerjee, A. V. (1992). A simple model of herd behavior. The Quarterly Journal of Economics 107(3), 797–817.

Battiston, P. and L. Stanca (2014). Boundedly rational opinion dynamics in directed social networks: theory and experimental evidence. University of Milan Bicocca Department of Economics, Management and Statistics Working Paper (267).

Beck, P. A., R. J. Dalton, S. Greene, and R. Huckfeldt (2002). The social calculus of voting: Interpersonal, media, and organizational influences on presidential choices. American Political Science Review 96(1), 57–73.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. Journal of Political Economy 100(5), 992–1026.

Chamley, C. (2004). Rational herds: Economic models of social learning. Cambridge University Press.

Chandrasekhar, A. G., H. Larreguy, and J. P. Xandri (2020). Testing models of social learning on networks: Evidence from two experiments. Econometrica 88(1), 1–32.

Corazzini, L., F. Pavesi, B. Petrovich, and L. Stanca (2012). Influential listeners: An experiment on persuasion bias in social networks. European Economic Review 56(6), 1276–1288.

DeGroot, M. H. (1974). Reaching a consensus. Journal of the American Statistical association 69(345), 118–121.

DeMarzo, P. M., D. Vayanos, and J. Zwiebel (2003). Persuasion bias, social influence, and unidimensional opinions. The Quarterly Journal of Economics 118(3), 909–968.

Duflo, E. and E. Saez (2003). The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment. The Quarterly Journal of Economics 118(3), 815–842.

Gale, D. and S. Kariv (2003). Bayesian learning in social networks. Games and Economic Behavior 45(2), 329–346.

Golub, B. and M. O. Jackson (2010). Naive learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics 2(1), 112–149.

Grimm, V. and F. Mengel (2020). Experiments on belief formation in networks. Journal of the European Economic Association 18(1), 49–82.

Meyer, C. D. and I. Stewart (2023). Matrix analysis and applied linear algebra. SIAM.

Montgomery, J. D. (1991). Social networks and labor-market outcomes: Toward an economic analysis. The American Economic Review 81(5), 1408–1418.

Mossel, E., A. Sly, and O. Tamuz (2015). Strategic learning and the topology of social networks. Econometrica 83(5), 1755–1794.

Mueller-Frank, M. (2013). A general framework for rational learning in social networks. Theoretical Economics 8(1), 1–40.

Mueller-Frank, M. (2014). Does one bayesian make a difference? Journal of Economic Theory 154, 423–452.

Trusov, M., R. E. Bucklin, and K. Pauwels (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. Journal of marketing 73(5), 90–102.

Appendix A

Endogenous Socialization with Costly Behavior

A.1 Solution of One Activity Example

In this example, we assume that agents have an increasing and concave utility function because there is no convex health cost. Since there is only a constant marginal cost of socializing, a concave utility function is needed to obtain an interior optimal solution. So we assume the utility function as $u_i = (s_i + \gamma d_i)^\alpha$ where $\alpha \in (0,1)$. The problem of a type 1 agent is

$$\max_{s_1,d_1}(s_1 + \gamma d_1)^\alpha - c.t_1$$

Using $s_i = t_{iA}q_{iA} + t_{iB}q_{iB}$ and $d_i = t_{iA}(1-q_{iA}) + t_{iB}(1-q_{iB})$, the optimization problem can be rewritten as:

$$\max_{t_1}(t_1 q_1 + \gamma t_1(1-q_1))^\alpha - ct_1$$

The optimization leads to the FOC:

$$[t_1]: \ \alpha((1-\gamma)q_1 + \gamma)^{(\alpha-1)}(q_1 + \gamma(1-q_1)) - c = 0$$

$$t_1^{\alpha-1} = \frac{c}{\alpha}(q_1(1-\gamma) + \gamma)^{-\alpha}$$

$$t_1 = (\frac{\alpha}{c})^{\frac{1}{1-\alpha}}(q_1(1-\gamma) + \gamma)^{\frac{\alpha}{1-\alpha}} \tag{A.1}$$

Similarly, from type 2's problem, FOC yields

$$t_2 = (\frac{\alpha}{c})^{\frac{1}{1-\alpha}}(q_2(1-\gamma)+\gamma)^{\frac{\alpha}{1-\alpha}} \tag{A.2}$$

We know the meeting probabilities are

$$\begin{aligned} q_1 &= \frac{N_1 t_1}{N_1 t_1 + N_2 t_2} \\ q_2 &= 1 - q_1 = \frac{N_2 t_2}{N_1 t_1 + N_2 t_2} \end{aligned} \tag{A.3}$$

Parametrizing $\alpha = 0.5$, $\gamma = 0.5$, $N_1 = 100$ and $N_2 = 200$, the equations (3), (4) and (5) are solved simultaneously to get the following solutions:

$$q_1 = 2 - \sqrt{3} \qquad q_2 = -1 + \sqrt{3}$$

$$t_1 = \frac{3-\sqrt{3}}{8c^2} \qquad t_2 = \frac{\sqrt{3}}{8c^2}$$

Notice that for the large group, it results in

$$H_2 = \frac{s_2}{s_2 + d_2} = \frac{q_2 t_2}{q_2 t_2 + (1-q_2)t_2} = q_2 = -1 + \sqrt{3} > 0.66 = \frac{200}{200+100} = w_2$$

which means the large group has positive homophily.

For the small group, it results in

$$H_1 = \frac{s_1}{s_1 + d_1} = \frac{q_1 t_1}{q_1 t_1 + (1-q_1)t_1} = q_1 = 2 - \sqrt{3} < 0.33 = \frac{100}{200+100} = w_1$$

which means the small group has negative homophily.

A.2 Solution of the Model

In this part, we consider all the possible situations that can arise as an equilibrium of the system. Since there are many cases to consider, we first classify them as the following, based on how integrated/segregated the groups are:

- *Perfectly-Segregated cases* where groups are perfectly segregated into two separate rooms.
- *Perfectly-Integrated cases* where groups spend all of their time together.

- *Semi-Integrated cases* where groups spend some part of their time together.

In this proof, we focus on each case individually and show that the only possible equilibria are those stated in Proposition 1.

A.2.1 Perfectly-Segregated Possible Equilibria

Let us consider the possible cases of equilibria where the groups are perfectly segregated. These are the cases where one group spends all of their time available doing one of the activities, whereas the other group spends all of their time doing the other activity.

These are the two possible cases of corner solutions

a) $t_{1A} = 1$ and $t_{2A} = 0$.

b) $t_{1A} = 0$ and $t_{2A} = 1$.

We show that none of these cases constitute an equilibrium, given that $h > 1$, because the type that spends all their time in the costly activity always wants to deviate from the choice of $t_{iA} = 1$.

Recall the optimization problem of type 1 agent is

$$\begin{aligned} \max_{t_{1A}} \quad & t_{1A}q_{1A} + (1 - t_{1A})q_{1B} + \gamma(t_{1A}(1 - q_{1A}) + (1 - t_{1A})(1 - q_{1B})) - \tfrac{h}{2}t_{1A}^2 \\ \text{s.t.} \quad & t_{1A} - 1 \in [0, 1] \end{aligned}$$

The constraint $t_{1A} \in [0, 1]$ is actually divides into the two following constraints:

$$\begin{aligned} t_{1A} - 1 \leq 0 \\ -t_{1A} \leq 0 \end{aligned}$$

The Kuhn-Tucker conditions coming from the maximization problem of type 1 agents are

$$\frac{\partial \mathcal{L}}{\partial t_{1A}} = (q_{1A} - q_{1B})(1 - \gamma) - ht_{1A} - \lambda_1 + \lambda_2 = 0 \tag{A.4}$$

$$\begin{aligned} t_{1A} - 1 \leq 0 \quad & \lambda_1 \geq 0 \quad \lambda_1(t_{1A} - 1) = 0 \\ -t_{1A} \leq 0 \quad & \lambda_2 \geq 0 \quad \lambda_2(-t_{1A}) = 0 \end{aligned}$$

Consider when $\lambda_1 > 0$ and $\lambda_2 = 0$. Since $\lambda_1 > 0$, $t_{1A} = 1$ should be true. Using $t_{1A} = 1$ in (6):

$$(q_{1A} - q_{1B})(1 - \gamma) - h = \lambda_1 > 0$$

should be true by assumption. But notice that, since $\gamma \in (0,1)$ and $-1 < (q_{1A} - q_{1B}) < 1$, $(q_{1A} - q_{1B})(1 - \gamma) - h$ is negative for sure if $h > 1$.

Hence, if $h > 1$, $t_{1A} = 1$ is never a solution.

The same argument can be used for type 2 and their choice of $t_{2A} = 1$, which means that, no matter what the meeting probabilities are, $t_{2A} = 1$ is never a solution for $h > 1$. So it is shown that the cases considered here can never be an equilibrium, assuming $h > 1$.

A.2.2 Perfectly-Integrated Possible Equilibria

Now we consider the cases in which both groups spend all of their time together, either in one of the activities or in both of the activities simultaneously.

These are the three possible cases where

a) $t_{1A} \in (0,1)$ and $t_{2A} \in (0,1)$.

b) $t_{1A} = 0$ and $t_{2A} = 0$.

c) $t_{1A} = 1$ and $t_{2A} = 1$.

In here, we show that only the case in the second item constitutes an equilibrium, which is the equilibrium (i) of Proposition 1.

a) Consider the first case where $t_{1A} \in (0,1)$ and $t_{2A} \in (0,1)$.

Notice that in order to have such an interior solution for both types of agents, both of the First-Order Conditions coming from the optimization problems of both types should be satisfied. The First-Order Condition of type 1 is

$$(q_{1A} - q_{1B})(1 - \gamma) = ht_{1A}$$

Symmetrically, the First-Order Condition of type 2 is

$$(q_{2A} - q_{2B})(1 - \gamma) = ht_{2A}$$

Because of the unbiased matching process, notice that

$$q_{2A} - q_{2B} = (1 - q_{1A}) - (1 - q_{1B}) = -(q_{1A} - q_{1B})$$

Then the First-Order Condition of type 2 becomes

$$-(q_{1A} - q_{1B})(1 - \gamma) = ht_{2A}$$

Combining the two of them, we get the optimality condition of

$$t_{2A} = -t_{1A}$$

which can not be since the choices of time should be non-negative. Hence it is not an equilibrium.

b) Consider the second case where $t_{1A} = 0$ and $t_{2A} = 0$.

For this to be an equilibrium, there should not exist any profitable deviation.

Without loss of generality, let us consider a possible deviation for type 1 from $t_{1A} = 0$ to some $t_{1A} = \epsilon > 0$. The change in her payoff is

$$\left[\epsilon q_{1A} + (1 - \epsilon)q_{1B} + \gamma(\epsilon(1 - q_{1A}) + (1 - \epsilon)(1 - q_{1B})) - \frac{h}{2}\epsilon^2\right] - [q_{1B} + \gamma(1 - q_{1B})]$$

reorganizing it

$$\epsilon(q_{1A} - q_{1B})(1 - \gamma) - \frac{h}{2}\epsilon^2 < 0$$

For this deviation to be non-profitable, the expression above should be negative. Since it is assumed that individuals' unilateral deviations will not affect the meeting probabilities, $q_{1A} = 0$. So the expression is always negative. Hence, there is no profitable deviation of type 1. The same arguments can be applied to type 2 as well. So this constitutes an equilibrium.

c) Consider the third case where $t_{1A} = 1$ and $t_{2A} = 1$.

Following the same arguments from the Kuhn-Tucker conditions in section 1 of Appendix B, both types of agents have incentives to deviate from spending all of their time in the costly activity room, assuming $h > 1$.

A.2.3 Semi-Integrated Possible Equilibria

Lastly, we consider the cases in which both groups spend some of their time together in one of the rooms, whereas the other room is occupied only by one type.

These are the four possible cases where

a) $t_{1A} \in (0,1)$ and $t_{2A} = 1$.

b) $t_{1A} = 1$ and $t_{2A} \in (0,1)$.

c) $t_{1A} \in (0,1)$ and $t_{2A} = 0$.

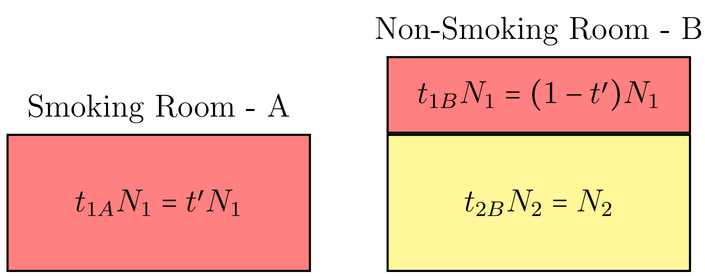
d) $t_{1A} = 0$ and $t_{2A} \in (0,1)$.

Here, we show that the cases in part (c) and (d) constitute equilibria, which are the equilibria (ii) and (iii) of Proposition 1.

For the cases in (a) and (b), it can be shown that they are not equilibrium using the same arguments from the Kuhn-Tucker conditions in section 1 of Appendix B.

c) Consider the third case where $t_{1A} \in (0,1)$ and $t_{2A} = 0$.

This situation can be visualized using the picture below which contains boxes with different colors showing the different types of people spending time in rooms A and B. The pink boxes show the stock of type 1 agents in each room, and the yellow box shows the stock of type 2 agents in room B.



Remember that the optimization problem of type 1 agent is

$$\max_{t_{1A}} \; s_1 + \gamma d_1 - \frac{1}{2}h^2 t_{1A} - ct_{1A} - c(1-t_{1A})$$

or equivalently

$$\max_{t_{1A}} \; t_{1A}q_{1A} + (1-t_{1A})q_{1B} + \gamma(t_{1A}(1-q_{1A}) + (1-t_{1A})(1-q_{1B})) - \frac{1}{2}h^2 t_{1A}$$

Notice that if we want the choice of $t_{1A} \in (0, 1)$ for a type 1 agent to constitute an equilibrium, we should ensure that it satisfies the optimality condition of type 1. However, for type 2, the similar optimality condition does not need to be satisfied as $t_{2A} = 0$ is a corner choice. As long as we ensure that a type 1 agent does not want to make a unilateral deviation, we would be safe to argue that it is an equilibrium.

The optimality condition for $t_{1A} = t' \in (0, 1)$ of type 1 comes from the First-Order Condition of the maximization problem of type 1:

$$(q_{1A} - q_{1B})(1 - \gamma) = ht_{1A}$$

Notice that the meeting probabilities in this situation are

$$q_{1A} = \frac{t'N_1}{t'N_1 + 0} = 1$$
$$q_{1B} = \frac{(1 - t')N_1}{(1 - t')N_1 + N_2}$$

Using these, the optimality condition becomes

$$(1 - \frac{(1 - t')N_1}{(1 - t')N_1 + N_2})(1 - \gamma) = ht'$$

When we solve this equation for t', we get two distinct solutions:

$$\frac{\sqrt{h}(N_1 + N_2) \pm \sqrt{4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2}}{2N_1\sqrt{h}}$$

Notice that the second solution $\frac{\sqrt{h}(N_1+N_2)+\sqrt{4N_1N_2(\gamma-1)+h(N_1+N_2)^2}}{2N_1\sqrt{h}}$ is always higher than 1 as $N_2 > N_1$, so it is not a possible solution. Therefore, we focus only on the other root of the system.

$$t' = \frac{\sqrt{h}(N_1 + N_2) - \sqrt{4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2}}{2N_1\sqrt{h}}$$

A few remarks on this solution:

- Assuming that $h > 1 - \gamma$, t' is a real number.

$$4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2 > 0$$

- t' is always positive since

$$h(N_1 + N_2)^2 > 4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2$$

$$0 > 4N_1N_2(\gamma - 1)$$

  always true as $\gamma - 1 < 0$.

- Assuming $h > 1 - \gamma$, $t' < 1$ holds.

$$\sqrt{h}(N_1 + N_2) - \sqrt{4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2} < 2N_1\sqrt{h}$$

$$\sqrt{h}(N_1 + N_2 - 2N_1) < \sqrt{4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2}$$

$$\sqrt{h}(N_2 - N_1) < \sqrt{4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2}$$

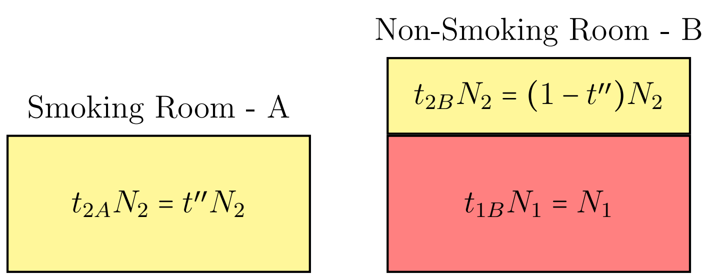
$$h(N_2 - N_1)^2 < 4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2$$

$$4(1 - \gamma)N_1N_2 < 4hN_1N_2$$

$$(1 - \gamma) < h$$

d) Consider the fourth case where $t_{1A} = 0$ and $t_{2A} = t'' \in (0, 1)$.

Again we visualize the situation with the picture below containing boxes with different colors showing the different types of people spending time in rooms A and B. The yellow boxes show the stock of type 2 agents in each room, and the pink box shows the stock of type 1 agents in room B.



Similar to the equilibrium above, the optimization problem of the type 2 agent is

$$\max_{t_{2A}} \; s_2 + \gamma d_2 - \frac{1}{2}h^2 t_{2A}$$

or equivalently

$$\max_{t_{2A}} t_{2A}q_{2A} + (1 - t_{2A})q_{2B} + \gamma(t_{2A}(1 - q_{2A}) + (1 - t_{2A})(1 - q_{2B})) - \frac{1}{2}h^2 t_{2A}$$

Again, the choice $t_{2A} \in (0, 1)$ for a type 2 agent to constitute an equilibrium should satisfy the optimality condition of type 2. However, for type 1, the similar optimality condition does not need to be satisfied as $t_{1A} = 0$ is a corner choice.

The optimality condition for $t_{2A} = t'' \in (0, 1)$ of type 2 comes from the First-Order Condition of the maximization problem of type 2:

$$(q_{2A} - q_{2B})(1 - \gamma) = ht_{2A}$$

Now, the meeting probabilities become

$$q_{2A} = \frac{t''N_2}{t''N_2 + 0} = 1$$
$$q_{2B} = \frac{(1 - t'')N_2}{(1 - t'')N_2 + N_1}$$

Using these, the optimality condition becomes

$$(1 - \frac{(1 - t'')N_2}{(1 - t'')N_2 + N_1})(1 - \gamma) = ht''$$

Solving this equation for t'', we get two distinct solutions:

$$\frac{\sqrt{h}(N_1 + N_2) \pm \sqrt{4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2}}{2N_2\sqrt{h}}$$

Notice that the second solution $\frac{\sqrt{h}(N_1+N_2)+\sqrt{4N_1N_2(\gamma-1)+h(N_1+N_2)^2}}{2N_2\sqrt{h}}$ is higher than 1 assuming that $h > 1 - \gamma$, so it is not a possible solution. Therefore, as we did in the previous part, we focus only on the other root of the system.

$$t'' = \frac{\sqrt{h}(N_1 + N_2) - \sqrt{4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2}}{2N_2\sqrt{h}}$$

A few remarks on this solution:

- Assuming that $h > 1 - \gamma$, t'' is a real number.

$$4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2 > 0$$

- t'' is always positive.

$$h(N_1 + N_2)^2 > 4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2$$

$$0 > 4N_1N_2(\gamma - 1)$$

always true as $\gamma - 1 < 0$.

- $t'' < 1$ is always true.

$$\sqrt{h}(N_1 + N_2) - \sqrt{4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2} < 2N_2\sqrt{h}$$

$$-\sqrt{4N_1N_2(\gamma - 1) + h(N_1 + N_2)^2} < \sqrt{h}(N_2 - N_1)$$

which is always true.

A.3 Comparative Statics

- Using the closed-from solutions of t' and t'', the partial derivative of t' with respect to N_1 is

$$\frac{\partial t'}{\partial N_1} = \frac{N_2((h - 2(1-\gamma))N_1 + hN_2 - \sqrt{h}\sqrt{h(N_1 + N_2)^2 - 4(1-\gamma)N_1N_2})}{2\sqrt{h}N_1^2\sqrt{h(N_1 + N_2)^2 - 4(1-\gamma)N_1N_2}}.$$

As the denominator is always positive, we focus only on the nominator:

$$(h - 2(1-\gamma))N_1 + hN_2 - \sqrt{h}\sqrt{h(N_1 + N_2)^2 - 4(1-\gamma)N_1N_2}) \lesseqgtr^{?} 0$$

$$(h - 2(1-\gamma))N_1 + hN_2 \lesseqgtr^{?} \sqrt{h}\sqrt{h(N_1 + N_2)^2 - 4(1-\gamma)N_1N_2})$$

Notice that both sides of the inequalities are positive when $N_2 > N_1$ and $h > 1 - \gamma$ is assumed. So taking the square of both sides does not change the sign of the inequality:

$$h^2(N_1+N_2)^2+4(1-\gamma)^2N_1^2-4(1-\gamma)hN_1(N_1+N_2) \lesseqgtr^{?} h^2(N_1+N_2)^2-4(1-\gamma)hN_1N_2$$

Simplifying it:

$$N_1((1-\gamma)-h)<0$$

is always true as $h>1-\gamma$ is assumed. So $\frac{\partial t'}{\partial N_1}<0$ is true.

- Following the same steps as above, it is also true that $\frac{\partial t''}{\partial N_2}<0$.

- The partial derivative of t' with respect to N_2 is

$$\frac{\partial t'}{\partial N_2}=\frac{1}{2N_1\sqrt{h}}(\sqrt{h}-\frac{2h(N_1+N_2)-4(1-\gamma)N_1}{2\sqrt{h(N_1+N_2)^2-4(1-\gamma)N_1N_2}})$$

$$=\frac{1}{2N_1\sqrt{h}}(\frac{2\sqrt{h(N_1+N_2)^2-4(1-\gamma)N_1N_2}-2h(N_1+N_2)+4(1-\gamma)N_1}{2\sqrt{h(N_1+N_2)^2-4(1-\gamma)N_1N_2}}).$$

Focusing only on the nominator, as the denominator is always positive,

$$\sqrt{h^2(N_1+N_2)^2-4h(1-\gamma)N_1N_2}-h(N_1+N_2)+2N_1(1-\gamma)\lesseqgtr^{?}0$$

$$\sqrt{h^2(N_1+N_2)^2-4h(1-\gamma)N_1N_2}\lesseqgtr^{?}h(N_1+N_2)-2N_1(1-\gamma).$$

Notice that taking the square of both sides of the inequality above does not change the direction of the inequality as both sides are always positive. More specifically, the right-hand side of the inequality above is always positive, as we assumed $h>1-\gamma$:

$$h>2(1-\gamma)-h$$

Multiplying the left-hand side by N_2 and the right-hand side by N_1, inequality still remains

$$hN_2>(2(1-\gamma)-h)N_1.$$

Hence

$$h(N_1+N_2)-2(1-\gamma)N_1>0.$$

So, we continue by taking the square of the former inequality:

$$h^2(N_1+N_2)^2-4h(1-\gamma)N_1N_2\lesseqgtr^{?}h^2(N_1+N_2)^2+4(1-\gamma)^2N_1^2-4h(1-\gamma)N_1(N_1+N_2)$$

$$4h(1-\gamma)N_1^2\lesseqgtr^{?}4(1-\gamma)^2N_1^2$$

Since it is assumed that $h>1-\gamma$, it is always true that $\frac{\partial t'}{\partial N_2}>0$.

- The partial derivative of t'' with respect to N_1 is

$$\frac{\partial t''}{\partial N_1} = \frac{1}{2N_2\sqrt{h}}(\sqrt{h} - \frac{2h(N_1+N_2) - 4(1-\gamma)N_2}{2\sqrt{h(N_1+N_2)^2 - 4(1-\gamma)N_1N_2}})$$

$$= \frac{1}{2N_2\sqrt{h}}(\frac{2\sqrt{h(N_1+N_2)^2 - 4(1-\gamma)N_1N_2} - 2h(N_1+N_2) + 4(1-\gamma)N_2}{2\sqrt{h(N_1+N_2)^2 - 4(1-\gamma)N_1N_2}}).$$

As we have shown above that

$$\sqrt{h^2(N_1+N_2)^2 - 4h(1-\gamma)N_1N_2} > h(N_1+N_2) - 2N_1(1-\gamma)$$

and because of the fact that $N_2 > N_1$:

$$\sqrt{h^2(N_1+N_2)^2 - 4h(1-\gamma)N_1N_2} > h(N_1+N_2) - 2N_1(1-\gamma) > h(N_1+N_2) - 2N_2(1-\gamma).$$

So, both the numerator and denominator of the partial derivative is positive, and $\frac{\partial t''}{\partial N_1} > 0$ is true.

- The partial derivative of t' with respect to γ is

$$\frac{\partial t'}{\partial \gamma} = -\frac{4N_1N_2}{4N_1\sqrt{h}\sqrt{h(N_1+N_2)^2 - 4(1-\gamma)N_1N_2}} < 0$$

which is always true, so $\frac{\partial t'}{\partial \gamma} < 0$.

- Following the same steps as above, it is also true that $\frac{\partial t''}{\partial \gamma} < 0$.

- The partial derivative of t' with respect to h is

$$\frac{\partial t'}{\partial h} = \frac{-(1-\gamma)N_1N_2}{N_1h\sqrt{h^2(N_1+N_2)^2 - 4(1-\gamma)hN_1N_2}} < 0$$

which is always negative as the denominator is positive. Hence, $\frac{\partial t'}{\partial h} < 0$.

- Symmetrically, it is true that $\frac{\partial t''}{\partial h} < 0$.

A.4 Pareto-Ranking

In this part, we Pareto-rank the three equilibria found in Appendix B, which are the following:

1. $t_{1A} = 0$ and $t_{2A} = 0$.

2. $t_{1A} = t' \in (0, 1)$ and $t_{2A} = 0$.

3. $t_{1A} = 0$ and $t_{2A} = t'' \in (0, 1)$.

the closed-form solutions for t' and t'' are indicated in Appendix B.

Let us denote the payoff of a type-i agent in the equilibrium described in section-k below for $k =\in \{1, 2, 3\}$ as $\mathcal{U}_i^k$.

A.4.1 Equilibrium where no one smokes

Notice that in the case of $t_{1A} = 0$ and $t_{2A} = 0$, there will not be socialization in Room A, and the meeting probability of agents in Room B is governed only by the population shares:

$$q_{1B} = \frac{t_{1B}N_1}{t_{1B}N_1 + t_{2B}N_2} = \frac{N_1}{N_1 + N_2}$$
$$q_{2B} = 1 - q_{1B} = \frac{t_{2B}N_2}{t_{1B}N_1 + t_{2B}N_2} = \frac{N_2}{N_1 + N_2}$$

Plugging the meeting probabilities into the payoff of agents, the resulting payoff of a type 1 agent:

$$\begin{aligned}
\mathcal{U}_1^1 &= s_1 + \gamma d_1 - \frac{h}{2}t_{1A}^2 \\
&= t_{1A}q_{1A} + (1 - t_{1A})q_{1B} + \gamma(t_{1A}(1 - q_{1A}) + (1 - t_{1A})(1 - q_{1B})) - \frac{h}{2}t_{1A}^2 \\
&= q_{1B} + \gamma(1 - q_{1B}) = \frac{N_1}{N_1 + N_2} + \gamma(1 - \frac{N_1}{N_1 + N_2}) = \frac{N_1 + \gamma N_2}{N_1 + N_2}
\end{aligned}$$

and for type 2 agent:

$$\begin{aligned}
\mathcal{U}_2^1 &= s_2 + \gamma d_2 - \frac{h}{2} t_{2A}^2 \\
&= t_{2A} q_{2A} + (1 - t_{2A}) q_{2B} + \gamma (t_{2A}(1 - q_{2A}) + (1 - t_{2A})(1 - q_{2B})) - \frac{h}{2} t_{2A}^2 \\
&= q_{2B} + \gamma(1 - q_{2B}) = \frac{N_2}{N_1 + N_2} + \gamma(1 - \frac{N_2}{N_1 + N_2}) = \frac{N_2 + \gamma N_1}{N_1 + N_2}
\end{aligned}$$

A.4.2 Equilibrium where only minority smokes

In the case where $t_{1A} = t' \in (0, 1)$ and $t_{2A} = 0$, only type 1 agents are in Room A, so their meeting probability there will be 1, whereas in Room B, there is a mixture of both types:

$$\begin{aligned}
q_{1A} &= 1 \\
q_{1B} &= \frac{(1 - t')N_1}{(1 - t')N_1 + N_2}
\end{aligned}$$

Plugging the meeting probabilities into the payoff of agents, the resulting payoff of a type 1 agent:

$$\begin{aligned}
\mathcal{U}_1^2 &= s_1 + \gamma d_1 - \frac{h}{2} t_{1A}^2 \\
&= t_{1A} q_{1A} + (1 - t_{1A}) q_{1B} + \gamma (t_{1A}(1 - q_{1A}) + (1 - t_{1A})(1 - q_{1B})) - \frac{h}{2} t_{1A}^2 \\
&= t' + (1 - t') \frac{(1 - t')N_1}{(1 - t')N_1 + N_2} + \gamma((1 - t')(1 - \frac{(1 - t')N_1}{(1 - t')N_1 + N_2})) - \frac{h}{2} t'^2 \\
&= \frac{(1 - t')N_1 + ((1 - \gamma)t' + \gamma)N_2}{(1 - t')N_1 + N_2} - \frac{h}{2} t'^2
\end{aligned}$$

and for type 2 agent:

$$\begin{aligned}
\mathcal{U}_2^2 &= s_2 + \gamma d_2 - \frac{h}{2} t_{2A}^2 \\
&= t_{2A} q_{2A} + (1 - t_{2A}) q_{2B} + \gamma (t_{2A}(1 - q_{2A}) + (1 - t_{2A})(1 - q_{2B})) - \frac{h}{2} t_{2A}^2 \\
&= q_{2B} + \gamma(1 - q_{2B}) = \frac{N_2}{(1 - t')N_1 + N_2} + \gamma(1 - \frac{N_2}{(1 - t')N_1 + N_2}) = \frac{N_2 + \gamma(1 - t')N_1}{(1 - t')N_1 + N_2}
\end{aligned}$$

where t' is the optimal time choice in the smoking room of type 1 agents derived in Appendix B.

A.4.3 Equilibrium where only the majority smokes

In the case where $t_{1A} = 0$ and $t_{2A} = t'' \in (0,1)$, only type 2 agents are in Room A, so their meeting probability there will be 1, whereas in Room B, there is a mixture of both types:

$$q_{1A} = 0$$
$$q_{1B} = \frac{N_1}{N_1 + (1 - t'')N_2}$$

Plugging the meeting probabilities into the payoff of agents, the resulting payoff of a type 1 agent:

$$\begin{aligned}
\mathcal{U}_1^3 &= s_1 + \gamma d_1 - \frac{h}{2} t_{1A}^2 \\
&= t_{1A} q_{1A} + (1 - t_{1A}) q_{1B} + \gamma (t_{1A}(1 - q_{1A}) + (1 - t_{1A})(1 - q_{1B})) - \frac{h}{2} t_{1A}^2 \\
&= q_{1B} + \gamma(1 - q_{1B}) = \frac{N_1}{N_1 + (1 - t'')N_2} + \gamma(1 - \frac{N_1}{N_1 + (1 - t'')N_2}) = \frac{N_1 + \gamma(1 - t'')N_2}{N_1 + (1 - t'')N_2}
\end{aligned}$$

and for type 2 agent:

$$\begin{aligned}
\mathcal{U}_2^3 &= s_2 + \gamma d_2 - \frac{h}{2} t_{2A}^2 \\
&= t_{2A} q_{2A} + (1 - t_{2A}) q_{2B} + \gamma (t_{2A}(1 - q_{2A}) + (1 - t_{2A})(1 - q_{2B})) - \frac{h}{2} t_{2A}^2 \\
&= t'' + (1 - t'') \frac{(1 - t'')N_2}{N_1 + (1 - t'')N_2} + \gamma((1 - t'')(1 - \frac{(1 - t'')N_2}{N_1 + (1 - t'')N_2})) - \frac{h}{2} t''^2 \\
&= \frac{(1 - t'')N_2 + ((1 - \gamma)t'' + \gamma)N_1}{N_1 + (1 - t'')N_2} - \frac{h}{2} t''^2
\end{aligned}$$

where t'' is the optimal time choice in the smoking room of type 2 agents derived in Appendix B.

Using the solutions of t' and t'', we can compare the utilities of each group in each case. Notice that for the minority group, the difference between their utility in the equilibria where only the minority smokes and only the majority smokes is:

$$\mathcal{U}_1^2 - \mathcal{U}_1^3 = \frac{(N_2 - 2N_1)(1 - \gamma)(h(N_1 + N_2) - \sqrt{h(4(\gamma - 1)N_1 N_2 + h(N_1 + N_2)^2)})}{2N_1(h(N_1 + N_2) + \sqrt{h(4(\gamma - 1)N_1 N_2 + h(N_1 + N_2)^2)})}$$

This difference is strictly positive if and only if $2N_1 < N_2$.

The same difference for the majority group is the following:

$$\mathcal{U}_2^2 - \mathcal{U}_2^3 = \frac{(2N_2 - N_1)(1-\gamma)(h(N_1+N_2) - \sqrt{h(4(\gamma-1)N_1N_2 + h(N_1+N_2)^2)})}{2N_2(h(N_1+N_2) + \sqrt{h(4(\gamma-1)N_1N_2 + h(N_1+N_2)^2)})}$$

which is always strictly positive.

Similarly, for the minority group, the difference in their utility in the equilibria where only the minority smokes and no one smokes is:

$$\mathcal{U}_1^2 - \mathcal{U}_1^1 = \frac{(N_2 - N_1)(1-\gamma)N_1(h(N_1+N_2) - \sqrt{h(4(\gamma-1)N_1N_2 + h(N_1+N_2)^2)})}{2N_1(h(N_1+N_2) + \sqrt{h(4(\gamma-1)N_1N_2 + h(N_1+N_2)^2)})}$$

This difference is always strictly positive.

The same difference for the majority group is the following:

$$\mathcal{U}_2^2 - \mathcal{U}_2^1 = \frac{N_2(1-\gamma)(h(N_1+N_2) - \sqrt{h(4(\gamma-1)N_1N_2 + h(N_1+N_2)^2)})}{(N_1+N_2)(h(N_1+N_2) + \sqrt{h(4(\gamma-1)N_1N_2 + h(N_1+N_2)^2)})}$$

which is always strictly positive as well.

A.5 Modification on Cost Structure I

Here we show that, given that an additional fixed cost component R in the total cost of activity A is within some interval, there exists a unilateral profitable deviation for type 2 agent from $t_{2A} = t'' \in (0,1)$ to $t_{2A} = 0$, but there is no unilateral profitable deviation for type 1 agent from $t_{1A} = t' \in (0,1)$ to $t_{1A} = 0$, given the closed form solutions of t' and t''.

So, here are the conditions we have to satisfy:

$$U_2(t_{2A} = 0) - (U_2(t_{2A} = t'') - R) > 0$$

meaning that deviating from $t_{2A} = t''$ to $t_{2A} = 0$ is profitable for type 2 in equilibrium (ii) is now profitable.

$$U_1(t_{1A} = 0) - (U_1(t_{1A} = t') - R) < 0$$

which says that the unilateral deviation from $t_{1A} = t' \in (0,1)$ to $t_{1A} = 0$ is not profitable for type 1 agent in equilibrium (i).

Here, the utility of playing $t_{2A} = 0$ when other type 2 agents play $t_{2A} = t''$ and type 1 agents play $t_{1A} = 0$ is the following:

$$s_2 + \gamma d_2 = t_{2A} q_{2A} + (1 - t_{2A}) q_{2B} + \gamma(t_{2A}(1 - q_{2A}) + (1 - t_{2A})(1 - q_{2A}))$$

$$= q_{2B} + \gamma(1 - q_{2B}) = \frac{(1 - t'')N_2}{N_1 + (1 - t'')N_2} + \gamma \frac{N_1}{(1 - t'')N_2 + N_1}$$

The payoff of type 2 agent from playing $t_{2A} = t''$ when other type 2 agents play $t_{2A} = t''$ and type 1 agents play $t_{1A} = 0$ is:

$$= t'' + (1 - t'')\left(\frac{(1 - t'')N_2 + \gamma N_1}{N_1 + (1 - t'')N_2}\right) - \frac{h}{2} t''^2 - R$$

In the same manner, the utility of playing $t_{1A} = 0$ when other type 1 agents play $t_{1A} = t'$ and type 2 agents play $t_{2A} = 0$ is the following:

$$s_1 + \gamma d_1 = t_{1A} q_{1A} + (1 - t_{1A}) q_{1B} + \gamma(t_{1A}(1 - q_{1A}) + (1 - t_{1A})(1 - q_{1A}))$$

$$= q_{1B} + \gamma(1 - q_{1B}) = \frac{(1 - t')N_1}{N_2 + (1 - t')N_1} + \gamma \frac{N_2}{(1 - t')N_1 + N_2}$$

The payoff of type 1 agent from playing $t_{1A} = t'$ when other type 1 agents play $t_{1A} = t'$ and type 2 agents play $t_{2A} = 0$ is:

$$= t' + (1 - t')\left(\frac{(1 - t')N_1 + \gamma N_2}{N_2 + (1 - t')N_1}\right) - \frac{h}{2} t'^2 - R$$

Using these expressions, the first inequality above becomes:

$$\frac{(1 - \gamma)N_1(h(N_1 + N_2) - \sqrt{h(4(\gamma - 1)N_1 N_2 + h(N_1 + N_2)^2)})}{2N_2(h(N_1 + N_2) + \sqrt{h(4(\gamma - 1)N_1 N_2 + h(N_1 + N_2)^2)})} < R$$

and the second inequality becomes:

$$R < \frac{(1 - \gamma)N_2(h(N_1 + N_2) - \sqrt{h(4(\gamma - 1)N_1 N_2 + h(N_1 + N_2)^2)})}{2N_1(h(N_1 + N_2) + \sqrt{h(4(\gamma - 1)N_1 N_2 + h(N_1 + N_2)^2)})}$$

As the ratios in both inequalities are always positive, such an interval exists for R such that equilibrium (ii) is not an equilibrium anymore but equilibrium (i) is still an equilibrium. Moreover, this interval, for the sake of brevity, is denoted as:

$$\frac{N_1}{N_2}G(N_1,N_2,h,\gamma) < R < \frac{N_2}{N_1}G(N_1,N_2,h,\gamma)$$

and

$$G(N_1,N_2,h,\gamma) = \frac{(1-\gamma)(h(N_1+N_2) - \sqrt{h(4(\gamma-1)N_1N_2 + h(N_1+N_2)^2)})}{2(h(N_1+N_2) + \sqrt{h(4(\gamma-1)N_1N_2 + h(N_1+N_2)^2)})}$$

If R is within this interval, then only equilibrium (i) survives (ignoring the pathological one - equilibrium (iii)). If R is smaller than the lower bound, then all three of the equilibria survive. If R is even higher than the upper bound, we only have the pathological equilibrium where no one chooses to do any costly activity.

A.6 Modification on Cost Structure II

Taking the total cost function as:

$$\text{Total Cost } (t_{iA}) = rt_{iA} + \frac{h}{2}t_{iA}^2$$

where $r, h > 0$,

- Notice that there is no fully-interior solution where both t_{iA}'s are in $(0,1)$. In any fully-interior solution, the optimality conditions for both types need to be satisfied:

  $$(q_{1A} - q_{1B})(1-\gamma) = r + ht_{1A}$$

  $$-(q_{1A} - q_{1B})(1-\gamma) = r + ht_{2A}$$

  Assume that there exists a t_{1A} that satisfies the condition above. Then

  $$\frac{(q_{1A} - q_{1B})(1-\gamma)}{h} = t_{1A} + \frac{r}{h} > 0$$

  But then

  $$\frac{-(q_{1A} - q_{1B})(1-\gamma)}{h} - \frac{r}{h} = t_{2A} < 0$$

  since $r, h > 0$ and $\frac{-(q_{1A}-q_{1B})(1-\gamma)}{h} > 0$. So there is no fully interior solution.

- From the Kuhn-Tucker conditions, for $t_{iA} = 1$ to be a solution, the following expression should be positive:

$$(q_{1A} - q_{1B})(1 - \gamma) - r - ht_{iA}$$

Then notice that if $r + h > 1 - \gamma$, $t_{iA} = 1$ is never a solution.

- Following the same rationale of proof of equilibria of the original model, $t_{1A} = 0$ and $t_{2A} = 0$ is an equilibrium.

- For the equilibria in which one type is not engaging in the costly activity at all, but the other does, following the same steps as in the original model, now we have two solutions:

  – $t_{1A} = t'_{new} = \frac{h(N_1+N_2)-N_1r-\sqrt{(N_1(h+r)+N_2h)^2-4h(1-\gamma)N_1N_2}}{2hN_1}$ and $t_{2A} = 0$.

  – $t_{1A} = 0$ and $t_{2A} = t''_{new} = \frac{h(N_1+N_2)-N_1r-\sqrt{(N_1(h+r)+N_2h)^2-4h(1-\gamma)N_1N_2}}{2hN_2}$.

  where both t'_{new} and t''_{new} are real numbers that are always smaller than 1, given that $h + r > 1 - \gamma$.After some calculations, it can easily be shown that t'_{new} is greater than 0 if and only if:

$$\frac{N_1}{N_2} < \frac{(1-\gamma) - r}{r}$$

  and $r < (1 - \gamma)$. Given that these conditions hold, t''_{new} is always less than 0, so t''_{new} is not a solution to the problem.

A.7 Regression Output

Table A.1: Dependent Variable: Smoking

	(1) Linear Prob. Model	(2) Linear Prob. Model	(3) Probit	(4) Probit
Population Share				
	(0.01)		(0.043)	
Minority dummy		0.0074		
		(0.005)		(0.024)
Parent Education				
	(0.003)	(0.003)	(0.013)	(0.013)
Age				
	(0.001)	(0.001)	(0.005)	(0.005)
Gender				
	(0.003)	(0.003)	(0.013)	(0.013)
School Dummy	*Yes*	*Yes*	*Yes*	*Yes*
Race Dummy	*Yes*	*Yes*	*Yes*	*Yes*
Adjusted R^2	0.06	0.06		
Number of Obs.	54255	54255	54255	54255

. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are in parentheses.

Table A.2: Dependent Variable: Drinking

	(1) Linear Prob. Model	(2) Linear Prob. Model	(3) Probit	(4) Probit
Population Share	−0.020.		−0.093*	
	(0.010)		(0.042)	
Minority dummy		0.002		0.017
		(0.732)		(0.023)
Parent Education	−0.012***	−0.012***	−0.054***	−0.054***
	(0.003)	(0.000)	(0.014)	(0.014)
Age	0.046***	0.046***	0.173***	0.173***
	(0.001)	(0.000)	(0.005)	(0.000)
Gender	0.043***	0.043***	0.166***	0.166***
	(0.003)	(0.000)	(0.013)	(0.013)
School Dummy	Yes	Yes	Yes	Yes
Race Dummy	Yes	Yes	Yes	Yes
Adjusted R^2	0.03	0.03		
Number of Obs.	54163	54163	54163	54163

. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are in parentheses.

Table A.3: Dependent Variable: Getting Drunk

	(1) Linear Prob. Model	(2) Linear Prob. Model	(3) Probit	(4) Probit
Population Share	−0.017*		−0.137**	
	(0.03)		(0.050)	
Minority dummy		0.0008		0.025
		(0.004)		(0.028)
Parent Education	−0.008**	−0.007**	−0.054**	−0.053**
	(0.004)	(0.002)	(0.016)	(0.001)
Age	0.031***	0.031***	0.172***	0.172***
	(0.000)	(0.001)	(0.006)	(0.006)
Gender	0.041***	0.041***	0.240***	0.241***
	(0.000)	(0.002)	(0.015)	(0.015)
School Dummy	*Yes*	*Yes*	*Yes*	*Yes*
Race Dummy	*Yes*	*Yes*	*Yes*	*Yes*
Adjusted R^2	0.06	0.06		
Number of Obs.	54163	54163	54163	54163

. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are in parentheses.

Table A.4: Dependent Variable: Racing

	(1) Linear Prob. Model	(2) Linear Prob. Model	(3) Probit	(4) Probit
Population Share	−0.035**		−0.117**	
	(0.001)		(0.038)	
Minority dummy		0.017**		0.055**
		(0.006)		(0.021)
Parent Education	0.008*	0.008*	0.029*	0.029*
	(0.003)	(0.003)	(0.013)	(0.001)
Age	−0.021***	−0.021***	−0.072***	−0.073***
	(0.001)	(0.001)	(0.005)	(0.005)
Gender	0.215***	0.216***	0.696***	0.696***
	(0.003)	(0.003)	(0.012)	(0.012)
School Dummy	Yes	Yes	Yes	Yes
Race Dummy	Yes	Yes	Yes	Yes
Adjusted R^2	0.06	0.06		
Number of Obs.	54163	54163	54163	54163

. $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Standard errors are in parentheses.

A.8 Different Thresholds of Costly Behavior

A.8.1 Weekly Threshold

	Dependent Variable: Smoking				Dependent Variable: Drinking			
	Linear Prob. Model		Probit		Linear Prob. Model		Probit	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Population Share	−0.040***	0.009	−0.253***	0.040	−0.018**	0.008	−0.126**	0.049
Minority Dummy	0.008	0.005	0.06**	0.025	0.002	0.004	0.024	0.027
	Dependent Variable: Getting Drunk				**Dependent Variable: Racing**			
	Linear Prob. Model		Probit		Linear Prob. Model		Probit	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Population Share	−0.016***	0.006	−0.175***	0.06	−0.027***	0.01	−0.104**	0.041
Minority Dummy	0.002	0.003	0.035	0.034	0.011**	0.005	0.044**	0.022

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A.8.2 Daily Threshold

	Dependent Variable: Smoking				Dependent Variable: Drinking			
	Linear Prob. Model		Probit		Linear Prob. Model		Probit	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Population Share	−0.040***	0.007	−0.353***	0.05	−0.009***	0.003	−0.255***	0.095
Minority Dummy	0.007*	0.004	0.09***	0.032	0.001	0.001	0.055	0.056

	Dependent Variable: Getting Drunk				Dependent Variable: Racing			
	Linear Prob. Model		Probit		Linear Prob. Model		Probit	
Variable	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Population Share	−0.008***	0.002	−0.283**	0.114	−0.019***	0.007	−0.129**	0.052
Minority Dummy	0.001	0.001	0.055	0.06	0.006*	0.003	0.039	0.028

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix B

Impact of Information Dissemination Strategies on Vaccine Decision: Empirical Evidence from Turkey

B.1 Survey Form

Here we show the English-translated version of the questionnaire form.

Survey

This questionnaire has been prepared to determine the importance of the criteria affecting the COVID-19 disease, prevention measures, information resources, vaccination status, and the decision to be vaccinated for individuals aged 25-45. This form contains questions about personal information, COVID-19 disease, preventive measures, COVID-19 vaccine status, and information resources about the disease and vaccine. Please read the questions carefully and give your most appropriate answer. Participation in the study is voluntary, and we will use your answers in our research without revealing your personal information or identity. Thank you for your interest and help.

1. **The Family Health Center and Family Health Unit you are registered in:**
   Town: ______
   Family Health Center Number: ______
   Family Health Unit Number: ______
2. **Age:** ______
3. **Sex:** ☐ Female ☐ Male
4. **Marital Status:** ☐ Married ☐ Single ☐ Widowed ☐ Divorced
5. **Level of Education:** ☐ None ☐ Primary ☐ Secondary ☐ University/Higher Education
6. **Employment Status:** ☐ Employed ☐ Unemployed
7. **Occupation:** ________
8. **Do you have any chronic illness?** ☐ No ☐ Yes (*Please indicate:* __________)
9. **Do you take any medication on a regular basis?** ☐ No ☐ Yes (*Please indicate:* __________)
10. **Have you ever had COVID-19?**

    ☐ No *(In this case, please skip to Question 13)*

    ☐ Yes
11. **When did you have COVID-19?** __/__/20__*(day/month/year)*
12. **How did you spend your recovery period from COVID-19?**

    ☐ At home, without medication

    ☐ At home, with medication

    ☐ In an hospital, but not in Intensive Care Unit(ICU)

    ☐ In Intensive Care Unit(ICU) in an hospital

    ☐ Other (*Please indicate:* ____________)
13. **Have you ever had the COVID-19 vaccine?**

    ☐ Yes *(In this case, please skip to Question 16)*

    ☐ No
14. **Are you planning on getting the COVID-19 vaccine?** ☐ Yes ☐ No
15. **Could you please mark your reason(s) for not getting the COVID-19 vaccine in the table below?** *Please answer this question if you have not been vaccinated against COVID-19. You can choose more than one option.*
    If you prefer to choose more than one option, can you specify in order of importance? *Write the reason you find the most important as 1 (one). Continue by typing 2 next to the second most important reason and 3 next to the third most important reason. Continuing in this way, write down the order of importance of all the items you choose to be relevant.*

Order of importance	Your reason(s) for not getting the COVID-19 vaccine
	I do not have enough information about the vaccine.
	I do not think the vaccine is effective.
	I think the vaccine has too many short-term side effects.
	I think the vaccine has severe long-term side effects.
	I am afraid of getting vaccinated.
	I do not trust people who will administer the vaccine.
	I do not think COVID-19 is a severe disease.
	COVID-19 vaccines are not yet tested adequately.
	The excipients in the vaccine will harm me.
	I had a lot of side effects from my previous vaccines.
	I am waiting for my turn to get the vaccine.
	I was busy and did not have the time to get the vaccine.
	I am waiting for my family members' turn to get the vaccine together.
	Other reason(s).(*Please indicate:* ______________)

16. Could you please mark your reason(s) for getting the COVID-19 vaccine in the table below?
Please answer this question if you have been vaccinated against COVID-19. You can choose more than one option.
If you prefer to choose more than one option, can you specify in order of importance?
Write the reason you find the most important as 1 (one). Continue by typing 2 next to the second most important reason and 3 next to the third most important reason. Continuing in this way, write down the order of importance of all the items you choose to be relevant.

Order of importance	Your reason(s) for getting the COVID-19 vaccine
	Not to get infected with COVID-19
	Not to be seriously ill due to COVID-19
	Not to die due to COVID-19
	Not to get infected with COVID-19 and infect my family and relatives
	I believe that the COVID-19 vaccine is effective.
	I think the side effects of the COVID-19 vaccine are not important.
	I believe I need to be vaccinated for the epidemic to stop.
	To be able to go to public areas safely (shopping malls, restaurants, cinema etc.)
	It is mandatory in my workplace.
	Because of my professional risk
	I have a chronic illness.
	Other reason(s).(*Please indicate:* ______________)

17. Please denote all the COVID-19 vaccines you have received, together with the vaccine brand and date.

			Brand			Date *(Month/Year)*
1^{st}dose	☐ Yes	☐ No	☐ Sinovac	☐ Biontech	☐ Turkovac	_/20__
2^{nd}dose	☐ Yes	☐ No	☐ Sinovac	☐ Biontech	☐ Turkovac	_/20__
3^{rd}dose	☐ Yes	☐ No	☐ Sinovac	☐ Biontech	☐ Turkovac	_/20__
4^{th}dose	☐ Yes	☐ No	☐ Sinovac	☐ Biontech	☐ Turkovac	_/20__
5^{th}dose	☐ Yes	☐ No	☐ Sinovac	☐ Biontech	☐ Turkovac	_/20__

18. Has anyone in your proximity (family, work or friends) had COVID-19 infection?

☐ Yes (*Please indicate your relation to them:* ___________)

☐ No

19. Has anyone in your proximity (family, work or friends) died due to COVID-19?

☐ Yes (*Please indicate your relation to them:* ___________)

☐ No

Can you mark the option that you think it fits you the best in the following cases?

	Totally agree	Agree	Not sure	Disagree	Totally disagree
20. I might get infected with COVID-19.					
21. I am afraid of getting infected with COVID-19.					
22. I am afraid of infecting family members with COVID-19.					

How often do you exhibit the protective behaviors in the following cases? Please mark the option that you think it fits you the best.

	Always	Often	Sometimes	Rarely	Never
23. I wear a mask outside to protect myself from COVID-19.					
24. I wash my hands to protect myself from COVID-19.					
25. I make sure that there is at least 1.5 meters between me and people outside.					
26. I avoid closed and public areas.					

27. Which of the following are the sources of information about the COVID-19 vaccine that you trust THE MOST? Can you denote the *three* of them that you think *the most* trustworthy by specifying the order?
Write the source you find the most trustworthy as 1 (one). Continue by typing 2 next to the second most trustworthy source and 3 next to the third most trustworthy source.

Order of importance	Information source(s)
	Web-pages of international organizations(such as World Health Organization(WHO))
	Official statements of scientists
	Information sources on the internet
	Social Media
	Newspapers/television
	Vaccine Companies
	Family members
	Close friends
	Neighbors
	Physicians/Doctors (other than your family physician)
	My family physician
	My family health nurse
	Pharmacists
	Other(s).(*Please indicate:* ______________)

28. Which health institution do you apply most frequently when you have a health problem? Can you denote the *three* of them that you visit *the most* frequently by specifying the order?
Write the institution you visit the most frequently as 1 (one). Continue by typing 2 next to the second most frequent institution and 3 next to the third most frequent institution.

Order of frequency	Health Institution(s)
	Family health center
	Public hospitals
	Public university hospitals
	Public polyclinics
	Private polyclinics
	Private hospitals
	Private practices
	Other(s).(*Please indicate:* ______________)

29. Which of the following are the sources of information about the COVID-19 vaccine that you DO NOT trust? Can you denote the *three* of them that you think *the least* trustworthy by specifying the order?
Write the source you find the least trustworthy as 1 (one). Continue by typing 2 next to the second least trustworthy source and 3 next to the third least trustworthy source.

Order of importance	Information source(s)
	Web-pages of international organizations(such as World Health Organization(WHO))
	Official statements of scientists
	Information sources on the internet
	Social Media
	Newspapers/television
	Vaccine Companies
	Family members
	Close friends
	Neighbors
	Physicians/Doctors (other than your family physician)
	My family physician
	My family health nurse
	Pharmacists
	Other(s).(*Please indicate:* ______________)

The Knowledge Rating Scale is a scale that shows the level of knowledge between 0 and 9 points (increasing from "0 - no information"to "9 - extensive information") to indicate the level of being informed. Please use this rating scale for your answers of questions 30, 31 and 32.

No Info → 0 1 2 3 4 5 6 7 8 9 ← Extensive Info

How much have you been informed by your family physician on the following subjects? *Please use the Knowledge Rating Scale above to mark the most appropriate answer for you.*

	No Info									**Extensive Info**
30. Information on the severity and course of the COVID-19 disease	0	1	2	3	4	5	6	7	8	9
31. Information on the effectiveness, application, side effects, safety, and availability of the COVID-19 vaccine	0	1	2	3	4	5	6	7	8	9
32. Information on where to get the COVID 19 vaccine	0	1	2	3	4	5	6	7	8	9

The Level of Impact on the Vaccination Decision Rating Scale is a scale that shows the level of influence between 0 and 9 points (increasing from "0 - has no impact"to "9 - has major impact") to indicate the level of impact on the vaccination decision. Please use this rating scale for your answers of questions 33, 34 and 35.

No Impact ↓ ... ↓ Major Impact

0 1 2 3 4 5 6 7 8 9

How much did the following information given by your family physician affect your decision to be vaccinated? *Please use the Level of Impact on the Vaccination Decision Rating Scale above to mark the most appropriate answer for you.*

	No Impact									**Major Impact**
33. Information on the severity and course of the COVID-19 disease	0	1	2	3	4	5	6	7	8	9
34. Information on the effectiveness, application, side effects, safety, and availability of the COVID-19 vaccine	0	1	2	3	4	5	6	7	8	9
35. Information on where to get the COVID 19 vaccine	0	1	2	3	4	5	6	7	8	9

36. How many times have you received information or consulted about the COVID-19 disease/vaccine from your family physician since the COVID-19 outbreak in 2020?: ________

37. How did you most often consult your family physician about COVID-19 disease/vaccine?

- ☐ Face-to-face meeting with the family physician
- ☐ Phone call
- ☐ Video call
- ☐ Information pamphlets
- ☐ Meeting with the nurse
- ☐ E-mail
- ☐ SMS/Whatsapp
- ☐ Other (*Please indicate:* ______________)

38. In your opinion, what was the most effective way to consult your family physician about COVID-19 disease/vaccine?

- ☐ Face-to-face meeting with the family physician
- ☐ Phone call
- ☐ Video call
- ☐ Information pamphlets
- ☐ Meeting with the nurse
- ☐ E-mail
- ☐ SMS/Whatsapp
- ☐ Other (*Please indicate:* ______________)

Thank you for your participation!

B.2 Descriptive Statistics

B.2.1 Health Care Units and Physicians

The sample consists of people who are assigned to eight different family physicians from five different family healthcare centers, and the assignment is done randomly by the government.

Variable	**Levels**	**n**	(%)
Health Center	balcova6	16	8.1
	buca35	52	26.4
	buca36	50	25.4
	gazi10	44	22.3
	gazi11	35	17.8
	Total	197	100.0
DoctorID	126	23	11.8
	127	27	13.8
	129	26	13.3
	130	24	12.3
	24	16	8.2
	31	33	16.9
	34	20	10.3
	35	26	13.3
	Total	195	100.0

Table B.1: Healthcare Centers and Doctors

Demographics

The sample consists of 197 people in total, of which %61.9 are women and %38.1 are men. People's age varies between 20 and 45. Also, apart from the smaller portion of only literate people (%3.5), the rest is almost equally distributed between middle school, high school, and university or above.

Variable	Levels	n	(%)
Sex	female	122	61.9
	male	75	38.1
	Total	197	100.0
Age Groups	20-25	11	5.7
	25-30	66	34.0
	30-35	48	24.7
	35-40	60	30.9
	41-45	9	4.6
	Total	194	100.0
Marital Status	single	49	25.0
	married	147	75.0
	Total	196	100.0
Education	literate	7	3.5
	middleschool	55	27.9
	highschool	67	34.0
	college	68	34.5
	Total	197	100.0
Employment	employed	108	55.1
	unemployed	88	44.9
	Total	196	100.0

Table B.2: Demographics

Medical Info

Variable	Levels	n	(%)
Chronic Illness	no	156	79.6
	yes	40	20.4
	Total	196	100.0
Medication	no	152	77.5
	yes	44	22.4
	Total	196	100.0
Had Covid	no	138	70.0
	yes	59	29.9
	Total	197	100.0

Table B.3: Medical Information

Variable	Levels	#times mentioned	Percentage based on responses (%)	Percentage based on cases (%)
Most trusted info	international organisations	91	18.6	50
source about the	scientists	93	19.1	51.1
COVID-19 vaccine	internet	31	6.4	17
	social media	12	2.5	6.6
	TV/newspaper	31	6.4	17
	vaccine companies	4	0.8	2.2
	family	14	2.9	7.7
	friends	7	1.4	3.8
	neighbors	4	0.8	2.2
	doctors	99	20.3	54.4
	my family physician	79	16.2	43.4
	my nurse	14	2.9	7.7
	pharmacists	5	1	2.7
	other	4	0.8	2.2
	Total	488	100.0	268.1
Least trusted info	international organisations	15	3.2	8.5
source about the	scientists	12	2.6	6.8
COVID-19 vaccine	internet	61	13	34.5
	social media	96	20.5	54.2
	TV/newspaper	68	14.5	38.4
	vaccine companies	61	13	34.5
	family	6	1.3	3.4
	friends	33	7.1	18.6
	neighbors	85	18.2	48
	doctors	9	1.9	5.1
	my family physician	7	1.5	4
	my nurse	3	0.6	1.7
	pharmacists	8	1.7	4.5
	other	4	0.9	2.3
	Total	468	100.0	264.4

Table B.4: Information Source Questionnaire

Variable	**Levels**	#times mentioned	Percentage based on responses (%)	Percentage based on cases (%)
Which health institution do you apply most frequently when you have a health problem?	family health center	170	33.7	87.6
	public hospitals	159	31.5	82
	public university hospitals	69	13.7	35.6
	public polyclinics	15	3	7.7
	private polyclinics	19	3.8	9.8
	private hospitals	60	11.9	30.9
	private practices	10	2	5.2
	other	2	0.4	1
	Total	504	100.0	259.8

Table B.5: Most frequently visited health institution

B.3 Regression Output

The following table shows the output of the regression made to assess the impact of information dissemination strategies on the information level of the patients about where/how to get the COVID-19 vaccine.

Table B.6: Results - Information level on where/how to get COVID-19 vaccine

Dependent variable: Info Level on where/how to get vaccine				
	(1)	(2)	(3)	(4)
Info strategy class (ref: Broadcasting)				
Personalized	3.056***	3.029***	2.587***	2.556***
	(0.595)	(0.603)	(0.838)	(0.868)
Doctor ID	Yes	Yes	Yes	Yes
Age (ref: age 20-25 years)				
25-30 years		0.787		0.461
		(0.772)		(0.869)
30-35 years		0.980		1.040
		(0.782)		(0.912)
35-40 years		0.754		0.744
		(0.777)		(0.880)
41-45 years		0.029		0.844
		(1.175)		(1.382)
Male		0.106		0.258
		(0.355)		(0.450)
Unemployed		0.009		0.053
		(0.362)		(0.451)
Chronic illness		0.775*		0.314
		(0.411)		(0.520)
Had covid-19			0.109	0.135
			(0.389)	(0.401)
Close death from covid-19			-0.651	-0.704
			(0.483)	(0.498)
Numb. of consultations			0.350***	0.337***
			(0.124)	(0.120)
Observations	171	167	129	126

Note: *p<0.1; **p<0.05; ***p<0.01*

Appendix C

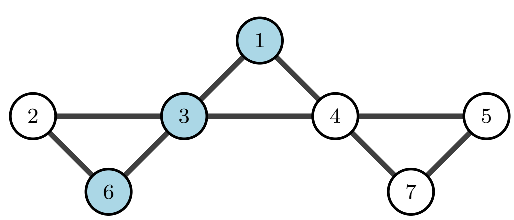
Social Learning and Degree of Coarseness in Communication

C.1 Modifications on the Example

C.1.1 Case I: Changing the Updating Rule

Let us assume that in Degroot model, the updating rule of agent i is to choose $a_{it} = 1$ if the posterior> $\frac{2}{3}$.

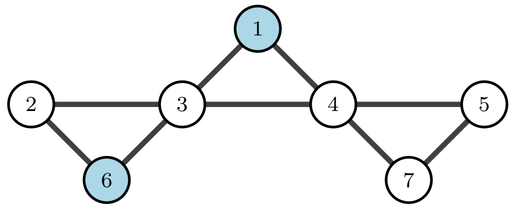
In $t = 1$, everyone plays their initial private signal:



For $t = 2$, agents update as the following:

$$\textit{For Mr.2: } \frac{0+1+1}{3} = \frac{2}{3} \textit{ so } a_2^2 = 0$$
$$\textit{For Mr.3: } \frac{0+0+1+1+1}{5} = \frac{3}{5} \textit{ so } a_3^2 = 0$$
$$\textit{For Mr.4: } \frac{0+1+1+0+0}{5} = \frac{2}{5} \textit{ so } a_4^2 = 0$$

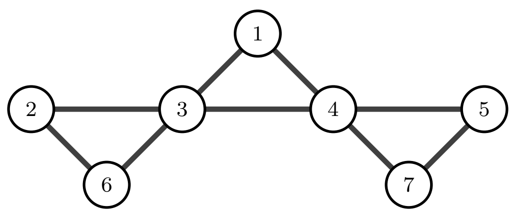
Hence the actions are:



For t=3, agents update as:

$$\textit{For Mr.2: } \frac{0+1+0}{3} = \frac{1}{3} \textit{ so } a_2^3 = 0$$
$$\textit{For Mr.1: } \frac{0+0+1}{3} = \frac{1}{3} \textit{ so } a_1^3 = 0$$

Hence the actions are:



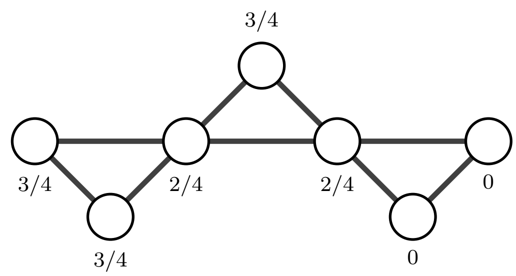
Notice that true learning happens! No agents get stuck in the wrong state.
Remark: Any posterior > $\frac{3}{5}$(which is the updating of Mr.3 at period 2) will end up with true learning.

C.1.2 Case II: Changing the Number of Possible Actions

Now let us assume that in DeGroot model, there are 5 possible states/actions: $\{0, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, 1\}$
In t=2, agents update as

$$\textit{For Mr.2: } \frac{0+1+1}{3} = \frac{2}{3} \textit{ is closest to } a_2^2 = \frac{3}{4}$$
$$\textit{For Mr.3: } \frac{0+0+1+1+1}{5} = \frac{3}{5} \textit{ is closest to } a_3^2 = \frac{2}{4}$$
$$\textit{For Mr.4: } \frac{0+1+1+0+0}{5} = \frac{2}{5} \textit{ is closest to } a_4^2 = \frac{2}{4}$$

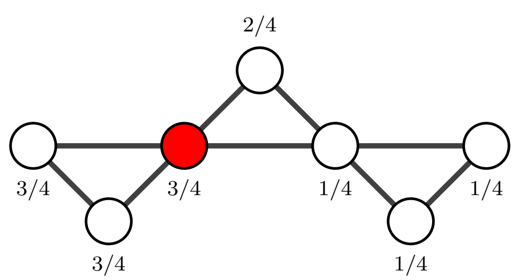
Hence the actions are:



For t=3, the updatings are as follows

$$\textit{For Mr.2: } \frac{3/4+3/4+2/4}{3} = \frac{2}{3} \textit{ is closest to } a_2^3 = \frac{3}{4}$$
$$\textit{For Mr.3: } \frac{2/4+2/4+3/4+3/4+3/4}{5} = \frac{13}{20} \textit{ is closest to } a_3^3 = \frac{3}{4}$$
$$\textit{For Mr.4: } \frac{3/4+2/4+2/4+0+0}{5} = \frac{7}{20} \textit{ is closest to } a_4^3 = \frac{1}{4}$$

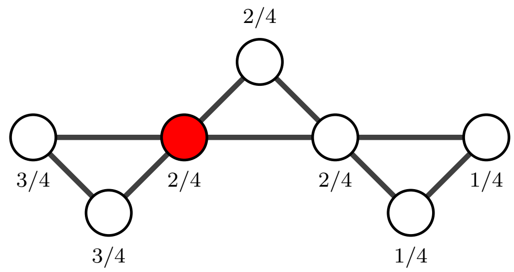
Hence the actions are:



In t=4, agents update as

$$\textit{For Mr.3: } \frac{2/4+1/4+3/4+3/4+3/4}{5} = \frac{12}{20} \textit{ is closest to } a_3^4 = \frac{2}{4}$$
$$\textit{For Mr.4: } \frac{3/4+2/4+1/4+0+0}{5} = \frac{6}{20} \textit{ is closest to } a_4^4 = \frac{1}{4}$$

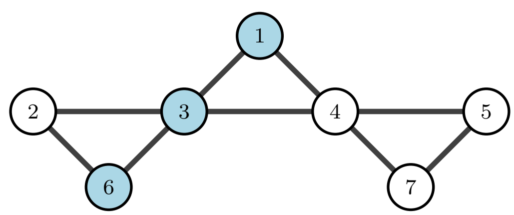
The actions are



As can be seen from the two graphs above, through the 3rd to 4th period, the opinion of Mr.3 has changed from $\frac{3}{4}$ to $\frac{2}{4}$ even though he is in the clan $\{2,3,6\}$ and the clan has the same action of $\frac{3}{4}$ in the 3rd period. ⇒ Violation of the proposition about "Stuckness of Clan".

C.1.3 Case III: Fine Communication

Assume that now agents can share their posterior beliefs, specifically agent i shares his updated belief about the probability that $\theta = 1$, $P_{it}(\theta = 1)$, and he takes the action 1 if $P_{it}(\theta = 1) > 1/2$.
Given that initial signals are distributed as before, agents take the following actions at $t = 1$:



their posterior beliefs are

$$
\begin{aligned}
&P_{11}(\theta = 1) = 1 \qquad && P_{51}(\theta = 1) = 0\\
&P_{21}(\theta = 1) = 0 \qquad && P_{61}(\theta = 1) = 1\\
&P_{31}(\theta = 1) = 1 \qquad && P_{71}(\theta = 1) = 0\\
&P_{41}(\theta = 1) = 0
\end{aligned}
$$

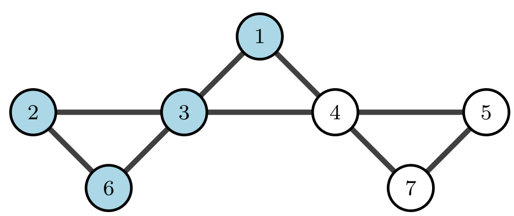
As everyone shares their posterior beliefs with their neighbors, they update their beliefs via DeGroot. For instance, as agent 1 is friends with agents 3 and 4:

$$P_{12}(\theta = 1) = \frac{\sum_{i \in \{1,3,4\}} P_i}{3} = \frac{1+1+0}{3} = \frac{2}{3}$$

and since this is greater than 1/2, agent 1 takes action 1, not 0. Similarly, the updated beliefs of others are

$$\begin{aligned} P_{22}(\theta = 1) &= \frac{2}{3} \qquad & P_{52}(\theta = 1) &= 0 \\ P_{32}(\theta = 1) &= \frac{3}{5} \qquad & P_{62}(\theta = 1) &= \frac{2}{3} \\ P_{42}(\theta = 1) &= \frac{2}{5} \qquad & P_{72}(\theta = 1) &= 0 \end{aligned}$$

So the actions at $t = 2$:



As the posterior beliefs are shared continuously, we can make use of the long-run properties of the Markov matrices (following Golub and Jackson (2010)), and calculate the long-run social influence vector to see what is the consensus belief of agents. It is

$$\begin{bmatrix} \frac{12}{100} & \frac{12}{100} & \frac{20}{100} & \frac{20}{100} & \frac{12}{100} & \frac{12}{100} & \frac{12}{100} \end{bmatrix}$$

So the long-run consensus belief of agents are

$$1.\frac{12}{100} + 0.\frac{12}{100} + 1.\frac{20}{100} + 0.\frac{20}{100} + 0.\frac{12}{100} + 1.\frac{12}{100} + 0.\frac{12}{100} = \frac{44}{100} < \frac{1}{2}$$

Hence for all i, $a_i^\infty = 0$. Then if we compare the actions in the 2nd period and in the ∞-time, we can see that the clan $\{2, 3, 6\}$ does not get stuck in the wrong state.

C.2 Proofs

C.2.1 Proof of Proposition 3.1

⇒ If a group C is a clan, then C is $\frac{1}{2}$-cohesive.

Assume that a group $C \in V$ is a clan. By definition of a clan, for all $i \in C$,

$$d_i(C) \geq d_i(V \backslash C) \tag{C.1}$$

where $d_i(C)$ is the degree of node $i \in C$ counted only among neighbors within C and $d_i(V\backslash C)$ is the degree of node i counted only among outside of group C.

If we add $d_i(C)$ to the both sides of inequality (1):

$$2d_i(C) \geq d_i(V\backslash C) + d_i(C)$$

Notice that $d_i(V\backslash C) + d_i(C)$ is the total degree of agent i.

$$2d_i(C) \geq d_i(V)$$
$$\frac{d_i(C)}{d_i(V)} \geq \frac{1}{2}$$

By definition,

$$d_i(C) = |N_i(V) \cap C| \text{:number of neighbors of i that lies in set C}$$
$$d_i(V) = |N_i(V)| \text{:number of i's neighbors}$$

Then the above inequality can be rewritten as

$$\frac{|N_i(V) \cap C|}{|N_i(V)|} \geq \frac{1}{2}$$

Since this holds for any $i \in C$, the following should also be true

$$\min_{i \in C} \frac{|N_i(V) \cap C|}{|N_i(V)|} \geq \frac{1}{2}$$

which is the definition of $\frac{1}{2}$-cohesiveness.

⇐ If C is $\frac{1}{2}$-cohesive, then C is a clan. Assume that a set of agents $C \in V$ is

$\frac{1}{2}$-cohesive. Then by definition

$$\min_{i \in C} \frac{|N_i(V) \cap C|}{|N_i(V)|} \geq \frac{1}{2} \tag{C.2}$$

Since the inequality holds for the agent i that gives the minimum value of the ratio, it should hold for all $i \in C$. We can pick an arbitrary $i \in C$, and the inequality (2) becomes

$$2|N_i(V) \cap C| \geq |N_i(V)|$$

Notice that since $N_i(V) \cap C$ and $N_i(V) \backslash C$ are two disjoint sets, we can write $|N_i(V)| = |N_i(V) \cap C| + |N_i(V) \backslash C|$. Then the above inequality becomes

$$\begin{aligned} 2|N_i(V) \cap C| &\geq |N_i(V) \cap C| + |N_i(V) \backslash C| \\ |N_i(V) \cap C| &\geq |N_i(V) \backslash C| \end{aligned}$$

Notice that

$$\begin{aligned} |N_i(V) \cap C| &= d_i(C) \\ |N_i(V) \backslash C| &= d_i(V \backslash C) \end{aligned}$$

following the notation of the paper by Chandrasekhar et al. (2020). Then the last inequality becomes

$$d_i(C) \geq d_i(V \backslash C)$$

which is the definition of the clan. ■

C.2.2 Proof of Proposition 3.2

For an arbitrary number of actions/states k, let us define the possible action set

$$A = \{0, \frac{1}{k-1}, \ldots\ldots\ldots \frac{k-3}{k-1}, \frac{k-2}{k-1}, 1\}$$

where $\forall t$, $\forall i$, $a_i^t \in A$. Figure 1 shows the representation of k different actions/states on the line segment $[0, 1]$.

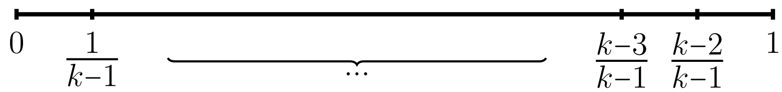


Figure C.1: Set of Possible Actions

Agent i holding the belief $b_i^t \in [x_{j-1}, x_j]$ (where x_{j-1} and x_j corresponds to any two consecutive elements of set A) choose his action a_t^i as

$$a_i^t = \left\{ \begin{array}{ll} x_j & \text{if } |b_i^t - x_j| > |b_i^t - x_{j-1}| \\ x_{j-1} & \text{if } |b_i^t - x_j| < |b_i^t - x_{j-1}| \end{array} \right\} \tag{C.3}$$

The interval $i^\star = \frac{1 - \frac{k-2}{k-1}}{2} = \frac{1}{2(k-1)}$ can be defined for a given k such that the agent chooses a specific action if and only if his belief lies within the $i^\star$ interval of that action.

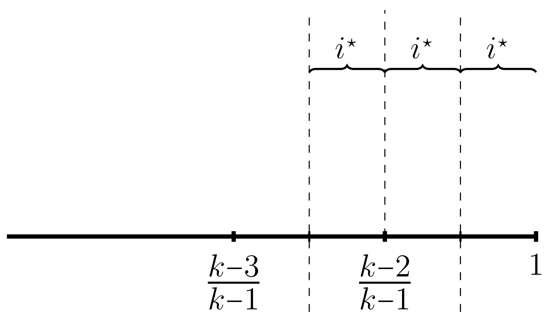


Figure C.2: Defining the interval

Using the definition of $i^\star$ and the rule for choosing the actions (3), we can restate the rule as follows:

Claim:

$$|b_i^t - x_j| < i^\star \iff a_i^t = x_j$$

Proof of Claim:

$\Rightarrow$ Assume $|b_i^t - x_j| < i^\star$ is true. That is

$$|b_i^t - x_j| < \frac{1}{2(k-1)}$$

<u>Case 1: For $b_t^i > x_j$</u>

$$b_i^t - x_j < \frac{1}{2(k-1)}$$

We know that for any consequent x_j and x_{j+1}, $x_{j+1} - x_j = \frac{1}{(k-1)}$ is true. Then substituting it into the above inequality:

$$b_i^t - (x_{j+1} - \frac{1}{(k-1)}) < \frac{1}{2(k-1)}$$

$$b_i^t - x_{j+1} < -\frac{1}{2(k-1)}$$

Since the right-hand side of the inequality is negative, the left-hand side should also be negative. That is, $b_i^t < x_{j+1}$. As $b_i^t > x_j$ and $b_i^t < x_{j+1}$ is shown, b_i^t is between x_j and x_{j+1} on the line segment $[0,1]$. Negating the inequality above:

$$-(b_i^t - x_{j+1}) > \frac{1}{2(k-1)}$$

Also since $b_i^t < x_{j+1}$, $|b_i^t - x_{j+1}| = -(b_i^t - x_{j+1})$ which means

$$|b_i^t - x_{j+1}| > \frac{1}{2(k-1)} \tag{C.4}$$

Combining the assumption that $|b_i^t - x_j| < \frac{1}{2(k-1)}$ and (4):

$$|b_i^t - x_{j+1}| > |b_i^t - x_j|$$

Then by the rule (3):

$$a_i^t = x_j$$

Case 2: For $b_i^t < x_j$

$$-(b_i^t - x_j) < \frac{1}{2(k-1)}$$

It is true for any consequent x_{j-1} and x_j, $x_j - x_{j-1} = \frac{1}{(k-1)}$ is true. Then substituting it into the inequality

$$-(b_i^t - \frac{1}{(k-1)} - x_{j-1}) < \frac{1}{2(k-1)}$$

$$-(b_i^t - x_{j-1}) < -\frac{1}{2(k-1)}$$

$$(b_i^t - x_{j-1}) > \frac{1}{2(k-1)}$$

Since the right-hand side of the inequality is positive, the left-hand side should also be positive. That is, $b_i^t > x_{j-1}$, which means $|b_i^t - x_{j-1}| = b_i^t - x_{j-1}$. Then we can restate the above inequality as

$$|b_i^t - x_{j-1}| > \frac{1}{2(k-1)} \tag{C.5}$$

Combining the assumption that $|b_i^t - x_j| < \frac{1}{2(k-1)}$ and (5):

$$|b_i^t - x_j| < |b_i^t - x_{j-1}|$$

Then by the rule (3):

$$a_i^t = x_j$$

$\Leftarrow$ Assume $a_i^t = x_j$ is true. Since actions are taken via rule (1), it should be true that agent i has a belief such that

$$|b_i^t - x_j| < |b_i^t - x_{j-1}| \tag{I}$$

$$|b_i^t - x_j| < |b_i^t - x_{j+1}| \tag{II}$$

where the belief $b_i^t \in [x_{j-1}, x_j]$.

Case 1: For $b_i^t > x_j$

From inequality (II) above

$$b_i^t - x_j < -(b_i^t - x_{j+1})$$

where $x_{j+1} - x_j = \frac{1}{(k-1)}$. Substituting inside

$$2b_i^t - x_j < \frac{1}{k-1} + x_j$$

$$b_i^t - x_j < \frac{1}{2(k-1)}$$

As $b_i^t > x_j$, it means $|b_i^t - x_j| < \frac{1}{2(k-1)}$ which is what we needed to show.

Case 2: For $b_i^t < x_j$

From inequality (I) above

$$-(b_i^t - x_j) < b_i^t - x_{j-1}$$

where $x_j - x_{j-1} = \frac{1}{(k-1)}$. Substituting inside

$$x_j < \frac{1}{k-1} - x_j + 2b_i^t$$

$$2x_j - 2b_i^t < \frac{1}{(k-1)}$$

$$-b_i^t + x_j < \frac{1}{2(k-1)}$$

As $b_i^t < x_j$, it means $|b_i^t - x_j| < \frac{1}{2(k-1)}$ which is what we needed to show to prove the claim. □

For an arbitrary agent i, let him have

- c number of clan member friends who all hold the opinion p.
- e number of friends outside of the clan holding some opinion $d_1, d_2, ...d_e$ such that all $d_k < p$, $\forall k \in \{1, ...e\}$.
- g number of friends outside of the clan holding some opinion $f_1, f_2, ...f_g$ such that all $f_l > p$, $\forall l \in \{1, ...g\}$.

where $c \geq e + g$.

Let us focus on what the next period belief of agent i be when he is observing the opinions p of c clan members, $d_1, d_2, ...d_e$ opinions of e and $f_1, f_2, ...f_g$ of g outside neighbors.

$$b^i = \frac{(c+1)p + \sum_{k\in\{1,...e\}} d_k + \sum_{l\in\{1,...g\}} f_l}{c+e+g+1}$$

Following the claim that we have proved above, it is true that to show that $a^i = p$ as $k \to \infty$, it is enough to show that $|b^i - p| \geq i^\star$ always holds.

$$\begin{aligned}|b^i - p| &= \left|\frac{(c+1).p + \sum_{k\in\{1,...e\}} d_k + \sum_{l\in\{1,...g\}} f_l}{c+e+g+1} - p\right| \\ &= \left|\frac{(\sum_{k\in\{1,...e\}} d_k + \sum_{l\in\{1,...g\}} f_l) - (e+g)p}{c+e+g+1}\right|\end{aligned}$$

Notice that this ratio does not depend on k at all, whereas

$$\lim_{k\to\infty} i^\star = \lim_{k\to\infty} \frac{1}{2(k-1)} = 0$$

Then it is clearly true that as $k \to \infty$

$$\left|\frac{(\sum_{k\in\{1,...e\}} d_k + \sum_{l\in\{1,...g\}} f_l) - (e+g).p}{c+e+g+1}\right| \geq \frac{1}{2(k-1)} \longrightarrow 0$$

surely holds since the left-hand side is always positive. ■

C.2.3 Proof of Proposition 3.3

Let there be k possible actions/states located on the unit line as:

$$A = \{0, \frac{1}{k-1}, ...\frac{k-3}{k-1}, \frac{k-2}{k-1}, 1\}$$

Suppose that there is a group of agents $S \in G$, such that S is $(1 - \frac{1}{2(k-1)})$-cohesive. It means:

$$\min_{i \in C} \frac{|N_i^\star(G) \cap C|}{|N_i^\star(G)|} \geq (1 - \frac{1}{2(k-1)}) \tag{C.6}$$

which means for every agent in S, at least $(1 - \frac{1}{2(k-1)})$ fraction of his friends belongs to group S as well.

Also, suppose that all agents in group S are DeGroot type, and they hold the same action a at some period of time t: $a_i^t = a$ for all $i \in S$.

I need to show that an agent with at least $(1 - \frac{1}{2(k-1)})$ fraction of his friends choosing the same action a will keep choosing the action a, even if the remaining fraction $\frac{1}{2(k-1)}$ of his friends choose some different action. Hence it is enough to show that he will choose the action a even when the action a and the others' actions are located at the furthest away possible from each other because in any other case, it would be easier for the updated belief to be close enough to the action held by the group.

In other words, we will consider the case in which the action of $(1 - \frac{1}{2(k-1)})$-cohesive group is located at 1, and all other friends of agent i choose the action located at 0. We will show that even in this extreme case, the next period action of agent i is 1 again which means he gets stuck on the $state = 1$, even if it may be the wrong state.

Let agent i have c number of friends belonging to the $(1 - \frac{1}{2(k-1)})$-cohesive group, all choosing $a^t = 1$ and d number of friends from outside of the group, all choosing $a^t = 0$. Notice that since $i \in S$, it is true for i that

$$\frac{|N_i^\star(G) \cap C|}{|N_i^\star(G)|} = \frac{c+1}{c+1+d} \geq (1 - \frac{1}{2(k-1)}) \tag{C.7}$$

Remember from the proof of the previous proposition, every agent $i \in G$ at each period of time chooses an action in a way that for all $x \in A$:

$$a_i^{t+1} = x \iff |b_i^{t+1} - x| < \frac{1}{2(k-1)}$$

Hence, in order to show that $a_i^{t+1} = 1$ it is enough to show that $|b_i^{t+1} - 1| < \frac{1}{2(k-1)}$ where b_i^{t+1} is the belief of agent i that he forms while observing the actions of c number of friends from set S, and d number of friends outside of set S.

$$|b_i^{t+1} - 1| = \left| \frac{(c+1).1 + 0.d}{c+d+1} - 1 \right| = \left| \frac{-d}{c+d+1} \right| = \frac{d}{c+d+1}$$

From the assumption of $(1 - \frac{1}{2(k-1)})$-cohesiveness, we have inequality (7):

$$\frac{c+1}{c+1+d} \geq \left(1 - \frac{1}{2(k-1)}\right)$$

If we subtract 1 from both sides

$$\frac{-d}{c+1+d} \geq \frac{-1}{2(k-1)}$$

$$\frac{d}{c+1+d} \leq \frac{1}{2(k-1)}$$

which was needed to be shown. ■