# Fighting Polarization with (Parental) Internalization*

Moti Michaeli† Jiabin Wu‡

August 27, 2021

**Abstract**

Growing polarization has been a rising concern in recent years all around the world. How can governments effectively fight it? In a dynamic model of inter-generational cultural transmission we show that polarization could be driven by the attempt of parents to instill extreme values in their children in anticipation of the pressure to conform that will be imposed on the children by their peers. However, this extremist tendency is mitigated if parents put a sufficiently large weight on their children's disutility from peer pressure and try to reduce it – rather than counterbalance it – by instilling in the children conformist rather than extreme values. Increasing awareness of parents to this peer pressure could therefore be a governmental tool for fighting polarization.

JEL-Code: D19, D91, J13, Z1.

Keywords: Inter-generational cultural transmission; Peer Pressure; Polarization; Norm Change.

*We thank Daniel Spiro for his constructive comments. Authors' email addresses: motimich@econ.haifa.ac.il; jwu5@uoregon.edu

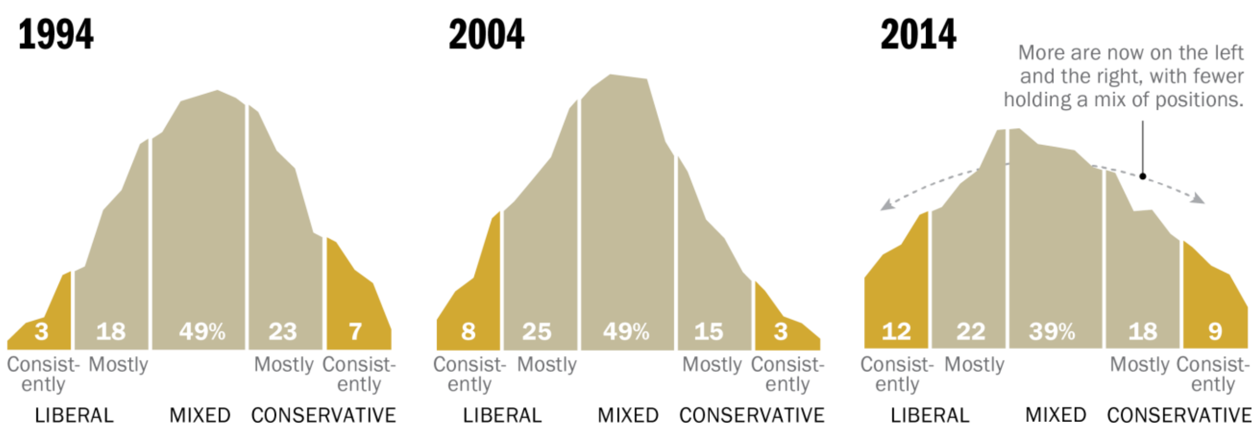
†University of Haifa

‡University of Oregon

# 1 Introduction

Growing polarization has been a rising concern in recent years all around the world. A report by Pew Research Center [2014] finds that in the United States, the proportion of the population holding extreme ideological views has been growing in the past two decades. From 1994 to 2014, the proportion of Americans in the tails of the ideological distribution—consistent liberals or consistent conservatives—has increased from 10% to 21%, and the center has shrunk from 49% to 39% (see Figure 1). This growing polarization is somewhat puzzling when considering peer effects, which tend to work in the direction of increasing rather than decreasing cohesion in society. However, recalling that values and views are shaped not only by peer effects but also by parental upbringing, the question arises: Could the upbringing of children be partly responsible for this increased polarization? And can we do anything to counterbalance this trend?

Figure 1: Polarization in the United State.




Source: Pew Research Center [2014]

The theory developed in the current paper provides an affirmative answer to both questions. We present a simple model in which two forces operate upon children and affect their behavior (actions): peer effects (or peer pressure) work in the direction of increasing integration and cohesion, while parental upbringing works in the direction of maintaining traditional views and values. Our model then shows that increased *parental internalization* of the peer pressure imposed on children can in fact mitigate polarization and lead to integration instead. By parental internalization of the peer pressure we refer to a disutility that is experienced by the parent if her child is the target of peer pressure (which could range from bullying and ostracizing to milder forms of social exclusion or simply social tensions).

The intuition is as follows. Parents want their children to make choices that are conducive to success in life, but they judge success from their own point of view (instead of their children's). Furthermore, parents may not be able to directly manipulate their children's decisions. Instead, they can shape their children's *preferences*, hoping to induce the children to choose the parents' desired actions.[1] When children's choices are independent of one another, a parent would simply want her child to have the exact preference trait as hers so that the child would make the optimal choice given her preference trait. However, if children are interacting with one another, especially when they are subject to pressure from their peers, a parent may have an incentive to instill in her child a preference trait that differs from her own. In particular, a parent might try to "overshoot" the preference she instills in her child in order to balance the conformist inclinations of the child which are prone to draw the child's action away from the parent's preference. This overshooting is what leads in our model to increased polarization from one generation to the next. However, if parents sufficiently internalize the peer pressure their children experience (which we interpret as a form of cost for the child), hence put a sufficiently large weight on reducing it, then this polarizing effect could be cancelled out and even reversed. The rationale is that in order to reduce the peer pressure on their children, parents need to instill in them views that are more conforming to those of other children. As a result, not only do the children experience less peer pressure, as intended by the parents, but also the views of the children's generation become less diverse, and polarization decreases. As hinted above, this result has a potential policy implication: in an attempt to reduce tensions in society, governments might wish to invest in campaigns encouraging parents to avoid instilling isolationist values in their children with the stated goal being to protect the children from social pressure.

## 1.1 Relation to the literature

Research shows that parents not only attempt to instill desirable traits in their children [Houser et al., 2016] but indeed often do manage to affect children's traits [Wilhelm et al.,

[1] Doepke and Zilibotti [2017] call this the "authoritative" parenting style, which lies in-between the "authoritarian" and the "permissive" parenting styles. An "authoritarian" parent restricts her child's choices. A "permissive" parent allows her child to make free choices according to the child's natural inclinations.

2008, Dohmen et al., 2012, Cappelen et al., 2020]. This paper is concerned with the effect of successful parental attempts to affect children's traits on the long-run distribution of traits in the society. In particular, our model shows that convergence to a unified trait is not inevitable as is claimed in, e.g., Cavalli-Sforza and Feldman [1973], Vaughan [2013] and Buechel et al. [2014], even when not using a probabilistic cultural transmission model as in Bisin and Verdier [2001] and Wu and Cheung [2018].[2] This is so because, anticipating the conformity-inducing effect of peer pressure, parents might choose to instill in their children values that are more extreme than their own, which leads to divergence of traits rather than to convergence.

Two closely related papers are Vaughan [2013] and Spiro [2020]. Vaughan [2013] is similar to our paper in taking into account the motive of children to conform to their peers. However, like the other papers mentioned above, Vaughan [2013] also concludes that convergence to one cultural trait is inevitable because he assumes that parents have control only over the weight of direct vs. oblique transmission but not on the transmitted values directly, an assumption that we find too restrictive. Furthermore, to some extent, Vaughan [2013] can be seen as a special case of our model where parents do not internalize the peer pressure imposed on their children. Spiro [2020] does get what he calls "perpetual extremizing", i.e. a sequence of generations who behave more extremely than their ancestor's bliss point, where in his model this result stems from parents caring about the behavior of multiple generations ahead (their grandchildren, grand-grandchildren and so on). However, and more importantly, in contrast to our own model (and just like Vaughan [2013] and the papers mentioned above), in Spiro's (2020) the society always converges to a unified trait in the long run.[3] While we do not find the assumption of caring for future generations implausible, the mechanism we explore in the current paper is somewhat more direct: parents simply adjust the instilled values to the anticipated peer pressure. Other papers that share with

[2] Also in Adriani et al. [2018] convergence is not inevitable, but there the mechanism is different: as an undesirable trait becomes rare, parents who hold this trait themselves stop hiding it, and this in turn increases the chance that this trait will be adapted by their children.

[3] The reason for this difference is quite intricate. In our model, actions are shaded versions of traits. Hence, if a parent cares about the child's action, she must instill an extreme trait in her child. In Spiro's (2020) model, in contrast, children's traits are shaded versions of (parental) actions. Hence, when a parent cares about the child's action (which will be similarly determined by the child as a tool for raising *her* child), and correctly anticipates the child to act as a parent like she does herself (namely to extremize her actions relative to her trait), she must instill a relatively moderate trait in the child. In the long run, this leads to the convergence of traits in the population.

Spiro [2020] the "role modelling" by parents, and share with both Spiro [2020] and our own paper the strategic choice of parents how to raise their children are Adriani and Sonderegger [2009, 2018], Adriani et al. [2018].

Two topics that are central in the current paper are polarization and the internalization (by parents) of peer pressure (on children). Polarization has been intensively studied in political science. See Layman et al. [2006] for a survey. Our paper is related to what is called the "popular polarization" in the literature, which views polarization as a decentralized phenomenon. See Hunter [1991], Frank [2004], and Fiorina et al. [2005] for book treatments. In economics, models based on Bayesian updating [Dixit and Weibull, 2007] and network theory [Marvela et al., 2011, Bolletta and Pin, 2020] are also offered to explain polarization. However, these researches do not account for the role of parenting – and of cultural transmission in general – in affecting polarization, as we do in this paper.

Internalization by parents of peer pressure imposed on their children is also well documented, as well as parental attempts to alleviate this pressure. For example, Liu-Farrer [2020] documents that in Japan, most immigrant children experience bullying at school, often because of their foreign-sounding names. In order to spare their children such experiences, some families apply for naturalization when their children enter school. Georgiou [2006] shows that parental attempts to prevent bullying and peer pressure on their children are sometimes even more extreme, up to a point where parents overprotect their children by preventing any interaction with other kids. On the other hand, when parents think peer pressure imposes a significantly negative effect on their children's behavior, they often take measures to curb it, in particular through instilling counter-balancing values in their children. Doepke and Zilibotti [2017] show that in recent decades, there is a trend toward more engaged parenting (in terms of time spending). That is, instead of restricting their children's choices, parents try to actively shape their children's preferences, hoping to induce them to take the parents' desired actions. Parents do so at least partly because their children are subject to peer effects that may divert them from choosing what the parents view as the "right thing to do". Indeed, Warr [1993] shows that the amount of time spent with family is capable of reducing and even eliminating peer influence. Similarly, Fernández-Villaverde et al. [2014] show that parents try to mold their daughters' attitudes toward premarital sex (by stigmatizing sex) to fight the tendency of young women to feel less shame from having an out-of-wedlock birth if their peers engage in premarital sex. Overall, the parental attitudes and actions documented in this paragraph motivate our way of modeling parental utility and actions.

The rest of the paper is organized as follows. Section 2 provides the model. Section 3 discusses the effect of adding the possibility of neighborhood segregation, which is often viewed as a determinant of polarization on the scale of the whole society.[4] Section 4 concludes.

# 2 The model

## 2.1 The intergenerational cultural transmission process

Consider a population of parents of unit mass. Each parent in the population has a trait from an unbounded set $T$.[5] The trait might be a political ideology, a measure of religiosity, a view on self reliance and so on. The population state is a distribution of traits over $T$ and thus is described by a probability measure $\mu$ over $T$. Adults reproduce asexually: Each parent bears one child and can influence the trait of this child. Then the child replaces her parent in the population.

Let $c^z \in T$ denote the trait of a child whose parent has trait $z \in T$. The child chooses an action, $x^z \in T$, to maximize the following objective function:

$$\max_{x^z} -(x^z - c^z)^2 - \int_{y \in T} (x^z - x^y)^2 \mu(dy). \tag{1}$$

That is, the child aims to minimize the sum of two kinds of disutility. First, the cognitive dissonance that results from choosing an action that differs from her trait $c^z$ (her bliss point). See Kuran and Sandholm [2008] for a discussion of the role that cognitive dissonance plays

[4] It is well documented that segregation can lead to polarization. For example, Johnson and Nazaryan [2019] find that classrooms today around the United States are as segregated as they were before the historic 1954 ruling in *Brown v. Board of Education*, and claim that one consequence of this trend of re-segregation is polarization of political attitudes. Relatedly, Alegre et al. [2008] find that assigning a single school to a single catchment area amplifies the patterns of residential segregation into the school-network social configuration. They argue that switching the policy to assigning multiple schools to each catchment area helps in reducing school segregation and polarization. Finally, Arora [2020] finds that segregation of some immigrant groups in the United States severely impedes the formation of the perception of common interests between the immigrants and the majority.

[5] An unbounded $T$ helps to simplify the analysis. Assuming that T is bounded would not change the main insights of the model. In Section 2.3, we provide an example with a bounded $T$.

in cultural evolution. Second, the peer pressure imposed by choosing an action that differs from the actions chosen by the other children (peer pressure in coordination games). See Michaeli and Spiro [2017] for a general model of peer pressure. Note that we assume that a child interacts with all other children globally. In Section 3, we introduce the possibility of segregation, where children interact only with peers from the same neighborhood.

We assume that the parent tries to instill a certain trait in her child while being aware that the child will also be influenced at the same time by the society at large, which is known as oblique socialization. More specifically, the trait of the child is a weighted average of the trait her parent tries to instill and the average trait of the other parents (other "role models" in society).

The parent's objective function is given as follows:

$$\max_{z'} -k(x^z - z)^2 - \alpha \int_{y \in T} (x^z - x^y)^2 \mu(dy) \tag{2}$$
$$\text{such that } c^z = \rho z' + (1-\rho) \int_{y \in T} y \mu(dy).$$

The constraint states that the child's trait $c^z$ is a linear combination of the trait that her parent chooses to instill ($z'$) and the societal average trait. The parameter $\rho$ measures the strength of oblique socialization. The smaller $\rho$ is, the more likely it is that the child will be influenced by the society at large.

The parent derives disutility if her child's action deviates from her own trait $z$ (instead of $c^z$). This element is aligned with the assumption of "imperfect empathy" of Bisin and Verdier [2001], according to which parents care about their children, yet are biased in their judgement (because they use their own traits to evaluate their children's behavior). The parameter $k > 0$ captures the level of (in)tolerance of the parent toward behaviors that deviate from her own bliss point.[6]

Consider the following two illustrative examples. As a first example, consider a vegan parent to a carnivorous child. The tolerance component captures how critical the parent is to the fact that the child eats meat – an intolerant vegan parent puts a large emphasis on this behavior being morally wrong from her point of view, while a tolerant parent believes

[6]Here we assume that the intolerance is increasing in the distance between one's own trait and the behavior she is evaluating, which is in line with, e.g., Guiso et al. [2006, 2009], Spolaore and Wacziarg [2009, 2013] and Wu and Cheung [2018].

in "live and let live". This component can also capture the extent to which the vegan parent thinks that a carnivorous *lifestyle* is unhealthy ("red meat will kill you"), hence not conducive for a successful life. The mirror image of this example is a carnivorous parent to a vegan child. The intolerance component captures the parent's distaste toward veganism ("these 'tree huggers'") or her concern about the health of her child ("where will you get your B12 vitamin from?"). As a second illustrative example, consider a woman whose gender-role attitude is intermediate in the society: she criticizes both women who think that a woman's only role in life is to be an obedient housewife and women who think that women are as free as men to pursue their careers and individual goals in life. Consequently, when considering her daughter's success in life, she might judge both a lack of an individualist career and a lack of "sufficient motherhood" (whatever this might be) as failures.

Our important and novel addition to the parental utility function is the second element in (2): we assume that the parent partly internalizes the peer pressure imposed on the child by her peers. This component is weighted by a parameter $\alpha > 0$, which captures the level of internalization. $\alpha$ is going to be central in our analysis. A large value of $\alpha$ means that the parent suffers when the child is subject to pressure (or potentially even bullying) by her peers. For example, a religious Muslim parent might wish her child to be a religious Muslim too, but if the child is raised in a secular Western society and the parent cares a lot about the inconvenience that the child might feel as an observant Muslim in such a society, the parent might wish to instill in the child a trait that is more reflective of the environment, namely to raise her child to be less observant of the Islam.[7] This is the intuition for why large $\alpha$ is going to be related to decreased polarization in society.

We solve the model by backward induction. Taking the first order condition of (1), we obtain

$$x^z = \frac{\int_{y \in T} x^y \mu(dy) + c^z}{2}, \tag{3}$$

which is the best response of a child with trait $c^z$ against other children. In words, the action of the child is an average of the trait instilled by the parent (and the other "role models")

[7] This example resonates with the empirical evidence in Liu-Farrer [2020] (about immigrant families in Japan applying for naturalization when their children enter school) that is mentioned in our literature review.

and the average action of the child's peers.[8] By solving all children's best response functions simultaneously, we obtain the equilibrium of the subgame played by all children given their traits.

**Lemma 1** *There exists a unique equilibrium in the subgame played by the children. In this equilibrium, the action taken by each child is the average of the trait instilled in her by her parent (and the other "role models") and the average trait of her peers:*[9]

$$x^z = \frac{\int_{y \in T} c^y \mu(dy) + c^z}{2}. \tag{4}$$

**Proof** See Appendix.

Next, we investigate parents' decisions. Plugging the children's equilibrium strategy profile into (2) and taking the first order condition, we obtain the best response of a parent with trait $z$ against other parents, which gives us

$$c^z = \frac{2kz + (\alpha - k) \int_{y \in T} c^y \mu(dy)}{k + \alpha}. \tag{5}$$

One can observe that the trait that the child eventually forms $c^z$ is a weighted average of the parent's own trait and the (expected) average trait of the other children. The weights put on each of these two elements depend on the value of the parameters $k$ and $\alpha$. A large $k$, capturing reluctance of the parent to raise a child whose behavior they disapprove, implies a large weight on the parent's trait. A large $\alpha$, on the other hand, implies a large weight on the average trait of the peers, because this would lower the peer pressure experienced by the child, a cost that is partly internalized by the parent.

Note that the optimal $c^z$ would not change if we assumed instead that the parent can directly choose $c^z$ rather than influence it through the choice of $z'$. Hence, we can abstract from oblique socialization, and will henceforth refer to $c^z$ as the trait that is (effectively) instilled by the parent. More importantly, from a parent's perspective, the expression of the

[8] We could of course turn this expression into a weighted average by putting different weights on the two components in the child's objective function. This would not change any of the results qualitatively.

[9] Note the difference from (3): the average action ($\int_{y \in T} x^y \mu(dy)$) is replaced by the average trait ($\int_{y \in T} c^y \mu(dy)$).

optimal $c^z$ reflects that the "competition" on influencing the child is between the parent and the child's peers rather than between the parent and other parents directly.

By solving all parents' best response functions simultaneously, we obtain the equilibrium strategies played by the parents given their traits, which leads to the following result.

**Lemma 2** *There exists a unique equilibrium of the game played between the parents and their children. In this equilibrium, the trait of each child is a weighted average of the parent's trait and the average trait of the parent's peers. More specifically,*[10]

$$c^z = \frac{2kz + (\alpha - k)\int_{y \in T} y\mu(dy)}{k + \alpha}. \tag{6}$$

**Proof** See Appendix.

Lemma 2 indicates that the child's trait in equilibrium boils down to a weighted average of the parent's trait and the trait of the parent's peers. Nevertheless, the weights are determined by the level of (in)tolerance $k$ and the level of internalization $\alpha$ rather than by the strength of oblique socialization $\rho$.

## 2.2 The dynamic model

We are now ready to put the parent-child interaction into an overlapping-generations dynamic model in order to study the evolution of traits (and actions) in society. From (4) to (6), observe that $\int_{y \in T} x^y \mu(dy) = \int_{y \in T} c^y \mu(dy) = \int_{y \in T} y\mu(dy)$. In other words, the average trait and the average action are constant and equal to each other across generations, only the variances are potentially evolving. This constant average, which we will denote by $\bar{y}$ for simplicity, forms the *social norm*. It is customary to identify the social norm with the average action in society (e.g. Glaeser and Scheinkman [2000], Özgür [2011], Michaeli and Spiro [2015]). Moreover, peer pressure is often assumed to be minimal when a person completely conforms to the norm (chooses $x^z = \bar{y}$ in our case), a property that holds also in our setup.[11]

[10]Note the difference from (5): the average trait among the children ($\int_{y \in T} c^y \mu(dy)$) is replaced by the average trait among the parents ($\int_{y \in T} y\mu(dy)$).

[11]This property holds for any convex peer pressure function, such as the quadratic function we have in (1), but does not hold when the peer pressure is concave – see Michaeli and Spiro [2017].

**Proposition 1** *The parameter $k$ and $\alpha$ determine the direction of the evolution of traits in society. In particular, for $z \neq \bar{y}$, $|c^z - \bar{y}| > |z - \bar{y}|$ if and only if $\alpha < k$.*

**Proof** See Appendix.

As described in Proposition 1, equation (6) implies that the relative value of $\alpha$ with respect to $k$ dictates the direction of evolution. When $\alpha < k$, the child's trait is further away from the social norm than the parent's trait is. When $\alpha = k$, the trait stays fixed across generations ($c^z = z$). When $\alpha > k$, the child's trait moves closer to the social norm. In other words, when parents do not internalize sufficiently the social cost embodied in the peer pressure ($\alpha < k$), they instill in their children traits that are more extreme than their own in order to counterbalance their children's tendency to conform to others' behavior due to peer pressure.[12] When parents do internalize the peer pressure sufficiently ($\alpha > k$), they have an incentive to instill in their children traits that are closer to the traits of their peers, thus reducing the peer pressure their children experience (or, from a different angle, allowing their children to grab the benefits of coordination). The following figures show the ordering of the key variables on a line for different values of $\alpha$ with respect to $k$ (for $z < \bar{y}$).[13]

Figure 2: Relative positions of the variables when $\alpha < k$

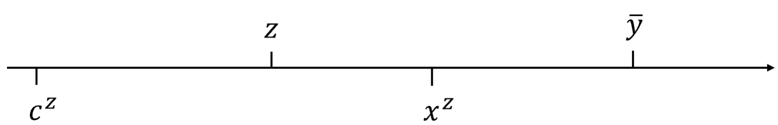
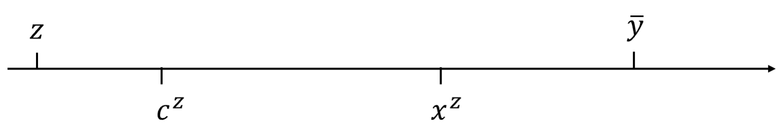



Figure 3: Relative positions of the variables when $\alpha > k$




Adding time indices, we can write down the cultural evolutionary dynamics: For any $\tilde{T} \subseteq T$, let $\hat{T} \subseteq T$ be the set such that for any $z \in \tilde{T}$, $c^z = \frac{2kz + (\alpha - k)\bar{y}}{k + \alpha} \in \hat{T}$; for any $c^z \in \hat{T}$,

[12]This observation shares some similarities with Spiro [2020].

[13]To understand how $x^z$ was located on the line, note that by plugging the equilibrium $c^z$ in (6) into (4), we get $x^z = \frac{kz + \alpha \int_{y \in T} y \mu(dy)}{k + \alpha}$.

$\frac{(k+\alpha)c^z+(k-\alpha)\bar{y}}{2k} \in \tilde{T}$ ($\tilde{T}$ and $\hat{T}$ are thus a pair in a parent-child bijection). Then

$$\mu_{t+1}(\hat{T}) = \mu_t(\tilde{T}). \tag{7}$$

It is straightforward to see that when $\alpha > k$, i.e., when parents sufficiently internalize the peer pressure, there is convergence to full conformity. In contrast, when $\alpha < k$, there is gradual (and endless) divergence of traits. The following proposition summarizes this result.

**Proposition 2** *The long term characterization of the dynamics is as follows:*

- ***Polarization**: When $\alpha < k$, there is constant divergence of traits: for any finite $\Delta$, there exist time $t_\Delta$ such that $\mu_t([\bar{y}-\Delta, \bar{y}+\Delta]\backslash\{\bar{y}\}) = 0$ for any $t > t_\Delta$, while $\mu_t(\{\bar{y}\}) = \mu_0(\{\bar{y}\})$.*
- ***Cloning**: When $\alpha = k$, the distribution of traits is constant, as each child is a perfect clone of her parent.*
- ***Integration**: When $\alpha > k$, there is convergence to the unique steady state $\mu(\{\bar{y}\}) = 1$. In this case $x^z = c^z = \bar{y}$ for everyone in society, i.e., everyone conforms to the norm in terms of both views and actions.*

**Proof** See Appendix.

Propositions 1 and 2 are informative when considering ways to fight polarization in society. While a state of full uniformity of traits and actions is not an ideal worth pursuing, raising $\alpha$ in a polarizing world seems warranted as it is predicted to mitigate the polarizing tendencies. A policy aimed at doing this could be, e.g., a governmental campaign that draws parents' attention to the social environment of their children and to the abundance of instances of mutual harassment among children on the background of conflicting views and values. This way, the micro-level behavior of parents (aimed at reducing the social cost their children bear) would contribute to solving the macro-level problem of polarization.[14]

In the next subsection, we provide an example of a simple special case where the end state is a complete bipolarization of society.

[14] Of course, encouraging children to bully other children who hold different views than their own could make a similar effect on polarization but this is obviously not a good policy to endorse.

## 2.3 An example of polarization

To consider a simple illustrative example, suppose that the trait is bounded in $[0, 1]$ and that, in the initial state ($t = 0$), the distribution of traits in society is uniform. For example, this could capture the political views of a pluralistic society about the proportion of a person's salary that should be given back to the community (in the form of taxes). Alternatively, the trait could represent the preferred level of adherence to one's religion (especially in religions with multiple levels of adherence, like Islam or Judaism). Suppose further that the internalization parameter, $\alpha$, is smaller than $k$. Then, by Proposition 2, our model predicts polarization. Moreover, since in this example the trait space is bounded, the long term steady state is a bipolar distribution of the trait.

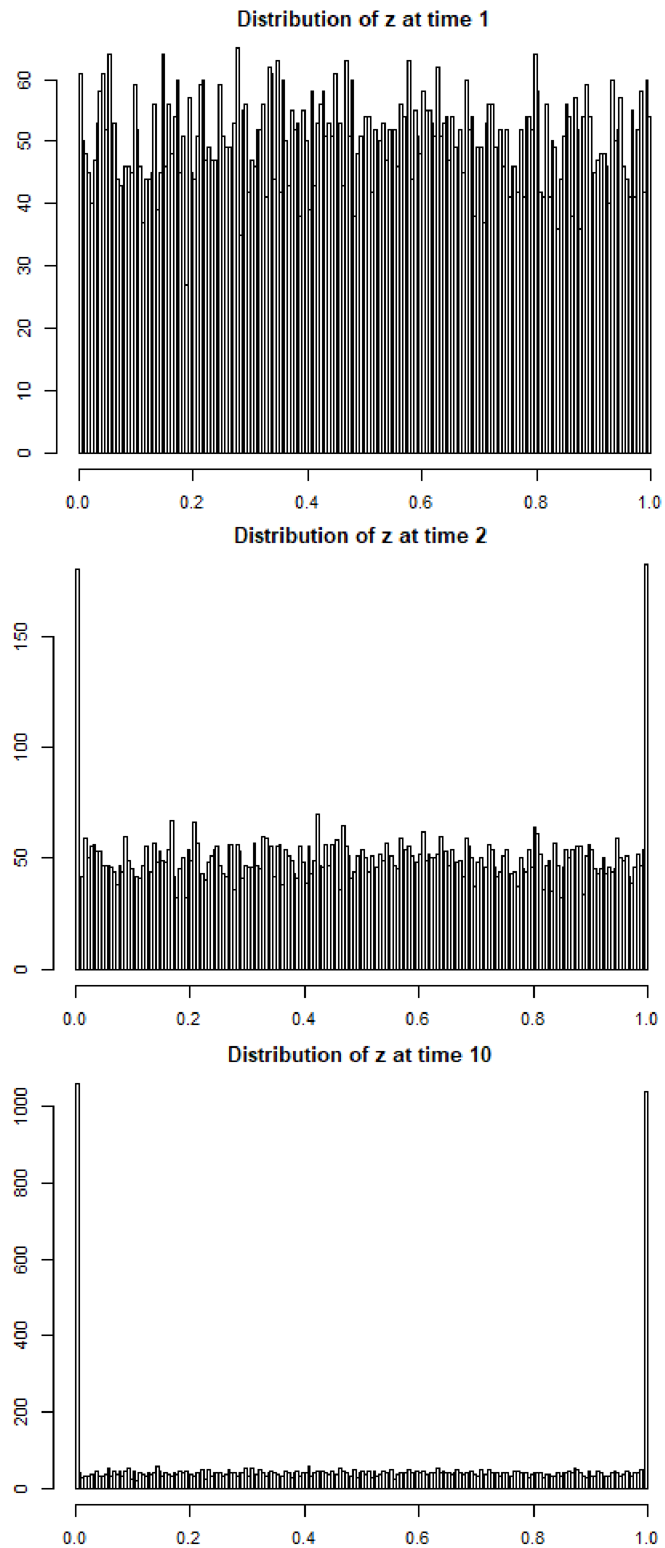
**Proposition 3** *Let $T \in [0, 1]$, the distribution of traits be uniform in the initial state, and $\alpha < k$. Then there is convergence to the unique steady state in which $\mu(\{\frac{1}{2}(1-\rho)\}) = \mu(\{\frac{1}{2}(1+\rho)\}) = \frac{1}{2}$.*

**Proof** See Appendix.

From Proposition 3, one can observe that the two polarized traits to which the dynamic eventually converges, $\frac{1}{2}(1-\rho)$ and $\frac{1}{2}(1+\rho)$, are closer to 0 and 1 as $\rho$ becomes larger. That is, as oblique socialization becomes weaker, the society is more polarized toward the extremes. Figure 4 depicts the dynamics for this example. We discretize the population, and set the population size to be 10000. We also set $k = 10$, $\alpha = 9.5$ and $\rho = 1$.[15] At time 0 (top schedule), the distribution of traits is (a random sampling from) a uniform distribution. Two time periods later (middle schedule), two peaks at the ends of the distribution start to emerge. Finally, by time 10 (bottom schedule), a large portion of society holds the two extreme traits, $z = 0$ and $z = 1$. In terms of the example at hand, this means that society divides into two hard-line extremist groups: the "ultra-libertarians" on the one hand, who wish to nullify taxes and government expenditures, and on the other hand the "communists", who deny any private earnings and private property. Or, in the example of religious adherence, society would divide into atheists/secular at one extreme, and religious zealots at the other extreme.

[15] The code in R is available upon request.

Figure 4: A simulation of polarization when the range of traits is bounded



# 3 Segregation

If the parents strongly internalize the peer pressure on their children, they might also be willing to take strong measures to fight this pressure. One of these measures could be to segregate themselves in a neighborhood with people whose views are similar to theirs. In this section we study this possibility and its effect on polarization.

Suppose there are two neighborhoods, $A$ and $B$, from which the parents choose one neighborhood to live in. Let $S_i$ denote the set of parents living in neighborhood $i$, for $i \in \{A, B\}$. $S_A \cup S_B = T$ and $S_A \cap S_B = \emptyset$. Assume that a parent can freely choose a neighborhood to live in to raise her child, where the child will only be subject to peer pressure within the neighborhood.

Let $c_i^z \in [0, 1]$ denote the trait that a child whose parent is living in neighborhood $i \in \{A, B\}$ and has trait $z \in S_i$ eventually forms. The child chooses an action, $x_i^z \in [0, 1]$ that maximizes the following objective function:

$$\max_{x_i^z} -(x_i^z - c_i^z)^2 - \int_{y \in S_i} (x_i^z - x_i^y)^2 \mu(dy)/\mu(S_i). \tag{8}$$

Taking the first order condition, we get

$$x_i^z = \frac{\int_{y \in S_i} x_i^y \mu(dy)/\mu(S_i) + c_i^z}{2}. \tag{9}$$

By integrating both sides over all $z \in S_i$, we get $\int_{y \in S_i} x_i^y \mu(dy) = \int_{y \in S_i} c_i^y \mu(dy)$. Hence, the equilibrium of the subgame played by the children in neighborhood $i \in \{A, B\}$ satisfies (compare with equation 4):

$$x_i^z = \frac{\int_{y \in S_i} c_i^y \mu(dy)/\mu(S_i) + c_i^z}{2}. \tag{10}$$

For simplicity, assume that the parent can directly choose the child's trait. The parent's objective function is given as follows:

$$\max_{i, c_i^z} -k(x_i^z - z)^2 - \alpha \int_{y \in S_i} (x_i^z - x_i^y)^2 \mu(dy)/\mu(S_i). \tag{11}$$

By plugging in the child's equilibrium action and taking the first-order condition, we get the parent's optimal choice of child's trait in each neighborhood:

$$c_i^z = \frac{2kz + (\alpha - k)\int_{y\in S_i} c_i^y \mu(dy)/\mu(S_i)}{k+\alpha}. \tag{12}$$

By integrating both sides over all $z \in S_i$, we get $\int_{y\in S_i} c_i^y \mu(dy) = \int_{y\in S_i} y\mu(dy)$. Hence, in equilibrium, we have (compare with equation 6):

$$c_i^z = \frac{2kz + (\alpha - k)\int_{y\in S_i} y\mu(dy)/\mu(S_i)}{k+\alpha}. \tag{13}$$

Let $\bar{y}_i = \int_{y\in S_i} y\mu(dy)/\mu(S_i)$ denote the average trait in neighborhood $i$. Observe that $\int_{y\in S_i} x_i^y \mu(dy)/\mu(S_i) = \int_{y\in S_i} c_i^y \mu(dy)/\mu(S_i) = \bar{y}_i$.

A parent's optimal choice of neighborhood depends on the comparison of her utility in each of the two neighborhoods given her respective choices of the instilled child's trait and the resulting action chosen by the child. The action chosen by the child of a parent with trait $z$ who lives in neighborhood $i \in \{A, B\}$ is given by

$$x_i^z = \frac{\bar{y}_i + \frac{2kz+(\alpha-k)\bar{y}_i}{k+\alpha}}{2} = \frac{kz + \alpha\bar{y}_i}{k+\alpha}. \tag{14}$$

Hence, the parent's utility in neighborhood $i$ is given by

$$\begin{aligned} U_z(i) &= -k\left(\frac{kz+\alpha\bar{y}_i}{k+\alpha} - z\right)^2 - \alpha\int_{y\in S_i}\left(\frac{kz+\alpha\bar{y}_i}{k+\alpha} - \frac{ky+\alpha\bar{y}_i}{k+\alpha}\right)^2\mu(dy)/\mu(S_i) \\ &= -\frac{k\alpha^2}{(k+\alpha)^2}(\bar{y}_i - z)^2 - \frac{k^2\alpha}{(k+\alpha)^2}\int_{y\in S_i}(z-y)^2\mu(dy)/\mu(S_i). \end{aligned} \tag{15}$$

First, observe that in any equilibrium, it is impossible to have either $\mu(S_A)$ or $\mu(S_B)$ be zero, because if there was an empty neighborhood, any parent would have an incentive to move into it, where there would be no peer pressure and she could hence choose her child's trait to exactly equal her own, thus maximizing her utility.

Second, suppose that $\bar{y}_A \neq \bar{y}_B$, and assume w.l.o.g. that $\bar{y}_B > \bar{y}_A$.[16] Then the equilibrium

[16] When $\bar{y}_A = \bar{y}_B$, then for any $z \in T$, $U_z(B) - U_z(A) = c$ for some constant $c$. Hence, to have $S_A$ and $S_B$ to be supported in equilibrium, we require $c = 0$. Constructing $S_A$ and $S_B$ such that $\bar{y}_A = \bar{y}_B$ and $c = 0$ is possible albeit difficult. Also, an equilibrium supporting $S_A$ and $S_B$ may be unstable against perturbations since every parent is indifferent between moving between neighborhoods.

division of the society into two neighborhoods takes a very specific form.

**Proposition 4** *Any equilibrium with $\bar{y}_B > \bar{y}_A$ has the following form: $S_A = (-\infty, z_{mid})$, $S_B = [z_{mid}, \infty)$, where $U_{z_{mid}}(A) = U_{z_{mid}}(B)$.*[17]

**Proof** See Appendix.

The proof of the proposition is based on the monotonicity of $U_z(B) - U_z(A)$ when $\bar{y}_A \neq \bar{y}_B$. This monotonicity implies that if $S_A$ and $S_B$ are indeed supported in equilibrium, then, as long as $\bar{y}_A \neq \bar{y}_B$, it is impossible to have $z < z' < z''$, such that $z, z'' \in S_A$ and $z' \in S_B$ or $z, z'' \in S_B$ and $z' \in S_A$. That is, there must exist some threshold $p$ dividing $S_A$ and $S_B$, such that $S_A = (-\infty, p)$ and $S_B = [p, \infty)$ or the other way around.

## 3.1 Segregation with parental internalization of peer pressure

Let us now turn our attention to a special case that sheds light on the dynamic unfolding of segregation and will be useful for studying the joint effect of segregation and parental internalization of peer pressure.

**Lemma 3** *Let $T \sim U[0,1]$. Then $S_A = [0, \frac{1}{2})$ and $S_B = [\frac{1}{2}, 1]$ is the unique equilibrium neighborhood division with $\bar{y}_B > \bar{y}_A$.*

**Proof** See Appendix.

In fact, if we consider the neighborhoods formation as a graduate dynamic process, this lemma implies that if the initial division is given by $S_A = [0, p)$, $S_B = [p, 1]$ for some $p \neq \frac{1}{2}$, the dynamic process would eventually converge to $S_A = [0, \frac{1}{2})$ and $S_B = [\frac{1}{2}, 1]$.

Suppose now that the parents sufficiently internalize the peer pressure, $\alpha > k$. Then, in the case considered under Lemma 3, $c_i^z \in S_i$ for any $z \in S_i$, $i \in \{A, B\}$ (see Figure 3). This in turn implies that in the next generation, when the children become parents themselves, staying in their childhood neighborhoods still constitutes an equilibrium (with $z_{mid}$ still being equal to $\frac{1}{2}$). The following corollary then immediately follows.

[17]For simplicity, we assume here that the distribution of traits is atomless at $z_{mid}$. If $\mu(\{z_{mid}\}) > 0$, then some of the $z_{mid}$ parents might live in neighborhood $A$, while the rest live in neighborhood $B$. Having this possibility does not change the flavor of the result.

**Corollary 1** *Let $T \sim U[0,1]$ and suppose $\alpha > k$. If initially $\bar{y}_B > \bar{y}_A$, then there is convergence to a unique steady state where $\mu(\bar{y}_A = \frac{1}{4}) = \mu(\bar{y}_B = \frac{3}{4}) = \frac{1}{2}$.*

The corollary applies the same logic of Proposition 2 (for the case of Integration) to each of the two neighborhoods that are depicted in Lemma 3. In words, it says that there will be dynamic converges to a state where all individuals in neighborhood $A$ have the trait $\frac{1}{4}$ and all individuals in neighborhood $B$ have the trait $\frac{3}{4}$. Hence, we have that segregation induces within-neighborhood homogenization and between-neighborhood differentiation, which is obviously more polarized than the completely homogeneous distribution of the Integration case in Proposition 2. However, note also that $\mu(\frac{1}{4}) = \mu(\frac{3}{4}) = \frac{1}{2}$ is *less* polarized than $\mu(0) = \mu(1) = \frac{1}{2}$, which is the distribution observed in Proposition 3 for the case of little internalization and no segregation (when $\rho = 1$ as assumed in the current section). To conclude, sufficient internalization of children's peer pressure *can* prevent extreme polarization even when segregation is possible. Nevertheless, segregation still serves as a severe impediment to the fight against polarization. Hence, eliminating segregation and increasing the awareness of parents to children's peer pressure should be taken into consideration by the policy makers at the same time.[18]

# 4 Conclusion

In this paper, we propose a dynamic intergenerational cultural-transmission model in which parents try to instill certain values in their children while incorporating the consideration that their children will face peer pressure to conform to the actions of their peers. We show that the dynamic can lead to polarization if parents do not sufficiently internalize this consideration and thus instill in their children extreme values. However, the trend toward polarization can be reversed if parents do sufficiently internalize this consideration and instead choose to instill more conformist values in their children. We thus point out that policies aiming to increase awareness of parents to children's peer pressure can be effective in fighting polarization.

[18]The case of $\alpha < k$ is analytically intractable and can produce multiple equilibria. For example, if $\alpha = 0$, it can be shown that any symmetric distribution of traits that has a mass of $\frac{1}{2} - m$ at each of the traits in $\{0,1\}$ while the rest is uniformly distributed on $[z, 1-z]$ with density equal to $\frac{2m}{1-2z}$ for $m = \frac{z}{0.5+z}$ forms a steady state, provided that $z \leq \frac{1}{4}$.

However, we also show that the possibility of segregation might substantially mitigate the effect that such internalization has on polarization: If parents segregate into different neighborhoods in an attempt to expose their children only to peers with traits that they approve, the internalization of peer pressure by the parents leads to trait homogeneity *within* neighborhoods but heterogeneity across neighborhoods. Still, this heterogeneity is shown to imply *less* polarization compared to the benchmark case where segregation is not possible but parents do not internalize the peer pressure on their children. In terms of policy implications, the grand lesson from the analysis of segregation is that in order to effectively fight polarization, it is not enough that parents internalize their children's peer pressure, but it is also important to eliminate segregation.

Instead of promoting policies that aim to increase awareness of parents to children's peer pressure, the government might try to promote a certain social norm that would serve as a driver of conformism in society, thereby reducing polarization. However, such an imposed norm may have a backlash effect. Hence, to study this possibility, one needs to extend the current model to incorporate the distuility from violating the governmental norm and to model the government as a player as well since it has its own agenda. We leave this interesting extension to future research.

# References

F. Adriani and S. Sonderegger. Why do parents socialize their children to behave prosocially? an information-based theory. *Journal of Public Economics*, 93(11-12):1119–1124, 2009.

F. Adriani and S. Sonderegger. Signaling about norms: Socialization under strategic uncertainty. *The Scandinavian Journal of Economics*, 120(3):685–716, 2018.

F. Adriani, J. A. Matheson, and S. Sonderegger. Teaching by example and induced beliefs in a model of cultural transmission. *Journal of Economic Behavior & Organization*, 145: 511–529, 2018.

M. A. Alegre, R. Benito, and I. Gonzalez. School segregation and polarization processes:

the influence of school zoning policies. *Profesorado. Revista de currículum y formación del profesorado*, 12, 2008.

M. Arora. Immigrant opposition in a changing national demographic. *Political Research Quarterly*, 73:340–351, 2020.

A. Bisin and T. Verdier. The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319, 2001.

U. Bolletta and P. Pin. Polarization when people choose their peers. Working Paper, 2020.

B. Buechel, T. Hellmann, and M. Pichler. The dynamics of continuous cultural traits in social networks. *Journal of Economic Theory*, 154:274–309, 2014.

A. Cappelen, J. List, A. Samek, and B. Tungodden. The effect of early-childhood education on social preferences. *Journal of Political Economy*, 128(7):2739–2758, 2020.

L. Cavalli-Sforza and M. Feldman. Cultural versus biological inheritance: phenotypic transmission from parents to children. *American Journal of Human Genetics*, 25:618–637, 1973.

A. K. Dixit and J. W. Weibull. Political polarization. *Proceedings of the National Academy of Science of the United States of America*, 104(18):7351–7356, 2007.

M. Doepke and F. Zilibotti. Parenting with style: Altruism and paternalism in intergenerational preference transmission. *Econometrica*, 85:1331–1371, 2017.

T. Dohmen, A. Falk, D. Huffman, and U. Sunde. The intergenerational transmission of risk and trust attitudes. *The Review of Economic Studies*, 79(2):645–677, 2012.

J. Fernández-Villaverde, J. Greenwood, and N. Guner. From shame to game in one hundred years: An economic model of the rise in premarital sex and its de-stigmatization. *Journal of the European Economic Association*, 12:25–61, 2014.

M. P. Fiorina, S. J. Abrams, and J. C. Pope. *Culture War? The Myth of a Polarized America*. Longman, 2005.

T. Frank. *What's the Matter with Kansas? How Conservatives Won the Heart of America.* Picador, 2004.

S. N. Georgiou. Bullying and victimization at school: The role of mothers. *British Journal of Educational Psychology*, 78:109–125, 2006.

E. L. Glaeser and J. Scheinkman. Non-market interactions. Technical report, National Bureau of Economic Research, 2000.

L. Guiso, P. Sapienza, and L. Zingales. Does culture affect economic outcomes? *Journal of Economic Perspectives*, 20:23–48, 2006.

L. Guiso, P. Sapienza, and L. Zingales. Cultural biases in economic exchange? *Quarterly Journal of Economics*, 124:1095–1131, 2009.

D. Houser, J. A. List, M. Piovesan, A. S. Samek, and J. Winter. On the origins of dishonesty: From parents to children. *European Economic Review*, 82:242–254, 2016.

J. D. Hunter. *Culture Wars: The Struggle to Define America.* Basic Books, 1991.

R. C. Johnson and A. Nazaryan. *Children of the Dream: Why School Integration Works.* New York, NY: Basic Books and Russell Sage Foundation Press, 2019.

T. Kuran and B. Sandholm. Cultural integration and its discontents. *Review of Economic Studies*, 75:201–228, 2008.

G. C. Layman, T. M. Carsey, and J. M. Horowitz. Party polarization in American politics: Characteristics, causes, and consequences. *Annual Review of Political Science*, 9:83–110, 2006.

G. Liu-Farrer. *Immigrant Japan: Mobility and Belonging in an Ethno-nationalist Society.* Cornell University Press, 2020.

S. A. Marvela, J. Kleinberg, R. D. Kleinberg, and S. H. Strogatz. Continuous-time model of structural balance. *Proceedings of the National Academy of Science of the United States of America*, 108(5):1771–1776, 2011.

M. Michaeli and D. Spiro. Norm conformity across societies. *Journal of public economics*, 132:51–65, 2015.

M. Michaeli and D. Spiro. From peer pressure to biased norms. *American Economic Journal: Microeconomics*, 9:152–216, 2017.

O. Özgür. Local interactions. In *Handbook of Social Economics*, volume 1, pages 587–644. Elsevier, 2011.

Pew Research Center. Political polarization in the american public, June 2014.

D. Spiro. Multigenerational transmission of culture. *Journal of Economic Theory*, 188: 105037, 2020.

R. Spolaore and R. Wacziarg. The diffusion of development. *Quarterly Journal of Economics*, 124:469–529, 2009.

R. Spolaore and R. Wacziarg. How deep are the roots of economic development? *Journal of Economic Literature*, 51:325–368, 2013.

D. Vaughan. Conformity and the transmission of continuous cultural traits. Mimeo, 2013.

M. Warr. Parents, peers, and delinquency. *Social Forces*, 72:247–264, 1993.

M. O. Wilhelm, E. Brown, P. M. Rooney, and R. Steinberg. The intergenerational transmission of generosity. *Journal of Public Economics*, 92(10-11):2146–2156, 2008.

J. Wu and M.-W. Cheung. On the probabilistic transmission of continuous cultural traits. *Journal of Economic Theory*, 174:300–323, February 2018.

# Appendix

## A Proofs

### A.1 Proof of Lemma 1

**Proof** Recall that the best response of the child with trait $c^z$ against other children is given by the first order condition in (3):

$$x^z = \frac{\int_{y \in T} x^y \mu(dy) + c^z}{2}.$$

By integrating both sides over all $z$, we have

$$\int_{z \in T} x^z \mu(dz) = \frac{\int_{y \in T} x^y \mu(dy) + \int_{z \in T} c^z \mu(dz)}{2},$$

which implies that $\int_{y \in T} x^y \mu(dy) = \int_{y \in T} c^y \mu(dy)$. Plugging this back into the first order condition, we obtain that

$$x^z = \frac{\int_{y \in T} c^y \mu(dy) + c^z}{2}.$$

### A.2 Proof of Lemma 2

**Proof** Plugging the children's equilibrium strategy profile into (2) and taking the first order condition, we obtain th best response of the parent with trait $z$ as

$$z' = \Big(\frac{2kz + (\alpha - k) \int_{y \in T} c^y \mu(dy)}{k + \alpha} - (1 - \rho) \int_{y \in T} y \mu(dy)\Big)/\rho, \tag{16}$$

which gives the optimal trait $c^z$ of the child as in (5):

$$c^z = \frac{2kz + (\alpha - k) \int_{y \in T} c^y \mu(dy)}{k + \alpha}.$$

By integrating both sides over all $z$, we have $\int_{y\in T} c^y \mu(dy) = \int_{y\in T} y\mu(dy)$. Hence, in equilibrium, we have

$$c^z = \frac{2kz + (\alpha - k)\int_{y\in T} y\mu(dy)}{k+\alpha}. \tag{17}$$

## A.3 Proof of Proposition 1

**Proof** By Lemma 2, we have

$$|c^z - \bar{y}| = \frac{2k}{k+\alpha}|z - \bar{y}|.$$

Hence, for $z \neq \bar{y}$, $|c^z - \bar{y}| > |z - \bar{y}|$ if and only if $\alpha < k$.

## A.4 Proof of Proposition 2

**Proof** For an individual with trait $z$ in the initial state, let $z_t$ denote the trait of her descendant in generation $t$. By Lemma 2, we have

$$|z_t - \bar{y}| = (\frac{2k}{k+\alpha})^t |z - \bar{y}|.$$

(1) When $\alpha < k$, if $z \neq \bar{y}$, $\lim_{t\to\infty} |z_t - \bar{y}| = \infty$. Hence, for any finite $\Delta > 0$, there exist time $t_\Delta$ such that $\mu_t([\bar{y} - \Delta, \bar{y} + \Delta]\backslash\{\bar{y}\}) = 0$ for any $t > t_\Delta$, while $\mu_t(\{\bar{y}\}) = \mu_0(\{\bar{y}\})$.

(2) When $\alpha = k$, $c^z = z$. Hence, $\mu_t = \mu_0$ for any $t > 0$.

(3) When $\alpha > k$, $\lim_{t\to\infty} |z_t - \bar{y}| = 0$. Hence, there is convergence to the unique state $\mu(\{\bar{y}\}) = 1$.

## A.5 Proof of Proposition 3

We first prove the following result:

**Lemma 4** *Let $T \in [0, 1]$. Then there exist $\hat{z}, \tilde{z} \in T$ with $\hat{z} \leq \tilde{z}$ such that, for any $z \in [0, \hat{z}]$, $z' = 0$ and $c^z = (1 - \rho)\bar{y}$; for any $z \in [\hat{z}, \tilde{z}]$, $z' \in (0, 1)$ and $c^z \in ((1 - \rho)\bar{y}, \rho + (1 - \rho)\bar{y})$; for* *any* $z \in [\tilde{z}, 1]$, $z' = 1$ *and* $c^z = \rho + (1-\rho)\bar{y}$. *The values of* $\hat{z}$ *and* $\tilde{z}$ *are given implicitly by*

$$\begin{cases} (1-\rho)\bar{y} = \dfrac{2k\hat{z} + (\alpha-k)\frac{2k\int_{y\in(\hat{z},\tilde{z})} y\mu(dy) + (\alpha+k)\left(\mu([0,\hat{z}])(1-\rho)\bar{y} + \mu([\tilde{z},1])(\rho+(1-\rho)\bar{y}\right)}{\alpha+k-\mu((\hat{z},\tilde{z}))(\alpha-k)}}{k+\alpha}, \\ \rho + (1-\rho)\bar{y} = \dfrac{2k\tilde{z} + (\alpha-k)\frac{2k\int_{y\in(\hat{z},\tilde{z})} y\mu(dy) + (\alpha+k)\left(\mu([0,\hat{z}])(1-\rho)\bar{y} + \mu([\tilde{z},1])(\rho+(1-\rho)\bar{y}\right)}{\alpha+k-\mu((\hat{z},\tilde{z}))(\alpha-k)}}{k+\alpha}. \end{cases} \tag{18}$$

**Proof** Recall that the best response of a parent with trait z against other parents is given by the first order condition:

$$z' = \big(c^z - (1-\rho)\bar{y}\big)/\rho, \tag{19}$$

where

$$c^z = \frac{2kz + (\alpha - k)\int_{y\in T} c^y \mu(dy)}{k+\alpha}. \tag{20}$$

When $\alpha < k$, the solution to (19) for sufficiently small $z$ is negative. However, since $T = [0,1]$, we get the corner solution $z' = 0$ for the parent's optimization problem. Similarly, for sufficiently large $z$, the solution to (19) is greater than 1, implying the corner solution $z' = 1$ for the parent's optimization problem. It thus follows that there exist $\hat{z}, \tilde{z} \in T$ with $\hat{z} \le \tilde{z}$ such that, for any $z \in [0, \hat{z}]$, $z' = 0$, which further implies by (2) that $c^z = (1-\rho)\bar{y}$; for any $z \in [\hat{z}, \tilde{z}]$, $z' \in (0,1)$, implying that $c^z \in ((1-\rho)\bar{y}, \rho + (1-\rho)\bar{y})$; and for any $z \in [\tilde{z}, 1]$, $z' = 1$, implying that $c^z = \rho + (1-\rho)\bar{y}$. By integrating $c^z$ over all $z$, we get

$$\begin{aligned} \int_{z\in T} c^z \mu(dz) \;&=\; \frac{2k\int_{z\in(\hat{z},\tilde{z})} z\mu(dz) + \mu((\hat{z},\tilde{z}))(\alpha-k)\int_{y\in T} c^y\mu(dy)}{\alpha+k} \\ &\quad + \mu([0,\hat{z}])(1-\rho)\bar{y} + \mu([\tilde{z},1])(\rho + (1-\rho)\bar{y}) \end{aligned}$$

which implies that

$$\int_{y\in T} c^y \mu(dy) = \frac{2k\int_{y\in(\hat{z},\tilde{z})} y\mu(dy) + (\alpha+k)\big(\mu([0,\hat{z}])(1-\rho)\bar{y} + \mu([\tilde{z},1])(\rho + (1-\rho)\bar{y}\big)}{\alpha + k - \mu((\hat{z},\tilde{z}))(\alpha-k)}. \tag{21}$$

Plugging this back into (20), we obtain that

$$c^z = \frac{2kz + (\alpha - k)\frac{2k\int_{y\in(\hat{z},\tilde{z})} y\mu(dy) + (\alpha+k)\left(\mu([0,\hat{z}])(1-\rho)\bar{y} + \mu([\tilde{z},1])(\rho+(1-\rho)\bar{y}\right)}{\alpha+k-\mu((\hat{z},\tilde{z}))(\alpha-k)}}{k+\alpha}, \text{ for } z \in (\hat{z},\tilde{z}).$$

One can then obtain the values of $\hat{z}$ and $\tilde{z}$ by plugging in $c^z = (1-\rho)\bar{y}$ and $c^z = \rho + (1-\rho)\bar{y}$ respectively.

### A.5.1 Proof of the Proposition

**Proof** Suppose that initially $z$ is uniformly distributed in $[0,1]$ (that is, $\mu_0 \sim U[0,1]$). By Lemma 4 there exist $\hat{z}, \tilde{z} \in T$ with $\hat{z} \le \tilde{z}$ such that, for any $z \in [0,\hat{z}]$, $z' = 0$ and $c^z = \frac{1}{2}(1-\rho)$; for any $z \in [\hat{z},\tilde{z}]$, $z' \in (0,1)$ and $c^z \in (\frac{1}{2}(1-\rho), \frac{1}{2}(1+\rho))$; for any $z \in [\tilde{z},1]$, $z' = 1$ and $c^z = \frac{1}{2}(1+\rho)$. Subtracting the two equations in (18) yields

$$2k(\tilde{z} - \hat{z}) = (k+\alpha)\rho.$$

Noting further that, due to symmetry, $\hat{z} = 1 - \tilde{z}$ we get

$$\hat{z} \;=\; \frac{1}{2} - \frac{(k+\alpha)\rho}{4k} > \frac{1}{2}(1-\rho), \quad \tilde{z} = \frac{1}{2} + \frac{(k+\alpha)\rho}{4k} < \frac{1}{2}(1+\rho).$$

The distribution of $c^z$ ($\mu_1$) is given by $\mu_1(\{\frac{1}{2}(1-\rho)\}) = \hat{z}$, $\mu_1(\{\frac{1}{2}(1+\rho)\}) = 1 - \tilde{z}$, and the corresponding density function on $(\frac{1}{2}(1-\rho), \frac{1}{2}(1+\rho))$ is given by $f_1(z) = \frac{\tilde{z}-\hat{z}}{\rho}$. Note that $\hat{z} = 1 - \tilde{z}$ and $\int_{z\in T} c^z \mu(dz) = \int_{z\in T} z\mu(dz) = \frac{1}{2}$. Hence, at time 1, the mean of the population stays at $\frac{1}{2}$, while the mass of individuals with trait $\frac{1}{2}(1-\rho)$ equals that with trait $\frac{1}{2}(1+\rho)$.

By induction, we have the same thresholds $\hat{z}$ and $\tilde{z}$ for each period and the mean stays at $\frac{1}{2}$ . Moreover, at time $t$, $\mu_t$ is given by $\mu_t(\{\frac{1}{2}(1-\rho)\}) = \Sigma_{k=1}^t \hat{z}(\tilde{z}-\hat{z})^{k-1} = \frac{1-(\tilde{z}-\hat{z})^t}{2}$, $\mu_t(\{\frac{1}{2}(1+\rho)\}) = \Sigma_{k=1}^t (1-\tilde{z})(\tilde{z}-\hat{z})^{k-1} = \frac{1-(\tilde{z}-\hat{z})^t}{2}$, and the corresponding density function on $(\frac{1}{2}(1-\rho), \frac{1}{2}(1+\rho))$ is given by $f_t(z) = (\tilde{z}-\hat{z})^t/\rho$. To see why, suppose $\mu_t$ is given as above. Then revisiting (21) we get

$$\int_{y\in T} c^y \mu(dy) = \frac{2k\int_{y\in(\hat{z},\tilde{z})} y\mu(dy) + (\alpha+k)\big(\mu([0,\hat{z}])\frac{1}{2}(1-\rho) + \mu([\tilde{z},1])\frac{1}{2}(1+\rho)\big)}{\alpha+k-\mu((\hat{z},\tilde{z}))(\alpha-k)}$$

$$
\begin{aligned}
&= \frac{2k\frac{1}{2}(\tilde{z}^2-\hat{z}^2)(\tilde{z}-\hat{z})^t/\rho}{\alpha+k-(\tilde{z}-\hat{z})^{t+1}(\alpha-k)\rho} \\
&+ \frac{(\alpha+k)((\hat{z}-\frac{1}{2}(1-\rho))(\tilde{z}-\hat{z})^t/\rho+\frac{1-(\tilde{z}-\hat{z})^t}{2})\frac{1}{2}(1-\rho)}{\alpha+k-(\tilde{z}-\hat{z})^{t+1}(\alpha-k)\rho} \\
&+ \frac{(\alpha+k)((\frac{1}{2}(1+\rho)-\tilde{z})(\tilde{z}-\hat{z})^t/\rho+\frac{1-(\tilde{z}-\hat{z})^t}{2})\frac{1}{2}(1+\rho)}{\alpha+k-(\tilde{z}-\hat{z})^{t+1}(\alpha-k)\rho}
\end{aligned}
$$

Using $\hat{z}+\tilde{z}=1$ to simplify the numerator of the first term, and using the fact that $\hat{z}-\frac{1}{2}(1-\rho)=\frac{1}{2}(1+\rho)-\tilde{z}$ for the numerator in the second term, the last expression equals

$$
\begin{aligned}
& \frac{k(\tilde{z}-\hat{z})^{t+1}/\rho+(\alpha+k)\big((\hat{z}-\frac{1}{2}(1-\rho))(\tilde{z}-\hat{z})^t/\rho+\frac{1-(\tilde{z}-\hat{z})^t}{2}\big)}{\alpha+k-(\tilde{z}-\hat{z})^{t+1}(\alpha-k)\rho} \\
= & \frac{k(\tilde{z}-\hat{z})^{t+1}/\rho+\frac{\alpha+k}{2}+(\alpha+k)\big((\hat{z}-\frac{1}{2}(1-\rho))/\rho-\frac{1}{2}\big)(\tilde{z}-\hat{z})^t}{\alpha+k-(\tilde{z}-\hat{z})^{t+1}(\alpha-k)/\rho} \\
= & \frac{k(\tilde{z}-\hat{z})^{t+1}/\rho+\frac{\alpha+k}{2}-(\alpha+k)(\frac{1}{2}-\hat{z})(\tilde{z}-\hat{z})^t/\rho}{\alpha+k-(\tilde{z}-\hat{z})^{t+1}(\alpha-k)/\rho} \\
= & \frac{k(\tilde{z}-\hat{z})^{t+1}/\rho+\frac{\alpha+k}{2}-\frac{\alpha+k}{2}(\tilde{z}-\hat{z})(\tilde{z}-\hat{z})^t/\rho}{\alpha+k-(\tilde{z}-\hat{z})^{t+1}(\alpha-k)/\rho} \text{ (using the fact that } \frac{1}{2}-\hat{z}=\frac{1}{2}(\tilde{z}-\hat{z})) \\
= & \frac{\frac{\alpha+k}{2}+(k-\frac{\alpha+k}{2})(\tilde{z}-\hat{z})^{t+1}/\rho}{\alpha+k-(\tilde{z}-\hat{z})^{t+1}(\alpha-k)/\rho} \\
= & \frac{\frac{\alpha+k}{2}-(\tilde{z}-\hat{z})^{t+1}(\alpha-k)/(2\rho)}{\alpha+k-(\tilde{z}-\hat{z})^{t+1}(\alpha-k)/\rho} = \frac{1}{2}.
\end{aligned}
$$

Hence, the two thresholds at time $t$ remain $\hat{z}$ and $\tilde{z}$. By adding the mass of children from parents with traits from $(\frac{1}{2}(1-\rho),\hat{z})$ to $\frac{1}{2}(1-\rho)$ and the mass of children from parents with traits from $(\tilde{z},\frac{1}{2}(1+\rho))$ to $\frac{1}{2}(1+\rho)$, we get that $\mu_{t+1}(\{\frac{1}{2}(1-\rho)\})=\Sigma_{k=1}^{t+1}\hat{z}(\tilde{z}-\hat{z})^k$, $\mu_{t+1}(\{\frac{1}{2}(1+\rho)\})=\Sigma_{k=1}^{t+1}(1-\tilde{z})(\tilde{z}-\hat{z})^k$, and consequently, the corresponding density function on $(\frac{1}{2}(1-\rho),\frac{1}{2}(1+\rho))$ is given by $f_{t+1}(z)=(\tilde{z}-\hat{z})^{t+1}/\rho$.

Observe that $\lim_{t\to\infty}\mu_t(\{\frac{1}{2}(1-\rho)\})=\lim_{t\to\infty}\Sigma_{k=1}^{t}\hat{z}(\tilde{z}-\hat{z})^{k-1}=\frac{\hat{z}}{1-(\tilde{z}-\hat{z})}=\frac{\hat{z}}{2\hat{z}}=\frac{1}{2}$. $\lim_{t\to\infty}\mu_t(\{\frac{1}{2}(1+\rho1\})=\lim_{t\to\infty}\Sigma_{k=1)}^{t}(1-\tilde{z})(\tilde{z}-\hat{z})^{k-1}=\frac{(1-\tilde{z})}{1-(\tilde{z}-\hat{z})}=\frac{1-\tilde{z}}{2(1-\tilde{z})}=\frac{1}{2}$.

## A.6 Proof of Proposition 4

**Proof** By taking the derivative of $U_z(B) - U_z(A)$ with respect to $z$, we get

$$\begin{aligned}
\frac{d(U_z(B) - U_z(A))}{dz} &= 2\frac{k\alpha^2}{(k+\alpha)^2}(\bar{y}_B - z) - 2\frac{k^2\alpha}{(k+\alpha)^2}\int_{y\in S_B}(z-y)\mu(dy)/\mu(S_B) \\
&\quad -2\frac{k\alpha^2}{(k+\alpha)^2}(\bar{y}_A - z) + 2\frac{k^2\alpha}{(k+\alpha)^2}\int_{y\in S_A}(z-y)\mu(dy)/\mu(S_A) \\
&= 2\frac{k\alpha^2}{(k+\alpha)^2}(\bar{y}_B - z) - 2\frac{k^2\alpha}{(k+\alpha)^2}(z - \bar{y}_B) \\
&\quad -2\frac{k\alpha^2}{(k+\alpha)^2}(\bar{y}_A - z) + 2\frac{k^2\alpha}{(k+\alpha)^2}(z - \bar{y}_A) \\
&= \frac{2k\alpha}{k+\alpha}(\bar{y}_B - \bar{y}_A) > 0. \qquad (22)
\end{aligned}$$

Hence, recalling that $S_A, S_B \neq \emptyset$, there exists a trait $z_{mid}$ s.t. $U_z(A) > U_z(B)$ for any $z \in S_A = (-\infty, z_{mid})$ and $U_z(B) > U_z(A)$ for any $z \in S_B \backslash \{z_{mid}\}$ where $S_B = [z_{mid}, \infty)$.

## A.7 Proof of Lemma 3

**Proof** Let $T \sim U[0,1]$ and suppose $\bar{y}_B > \bar{y}_A$. This implies by Proposition 4 that $S_A = [0,p)$ and $S_B = [p,1]$ for some $p \in (0,1)$. Now suppose by contradiction that $p > \frac{1}{2}$ (the proof is similar for $p < \frac{1}{2}$). In this case, $\bar{y}_A = \frac{p}{2}$, $\mu(S_A) = p$, and $\bar{y}_B = \frac{1+p}{2}$, $\mu(S_B) = 1-p$. By using Equation (15), we can calculate that

$$U_p(A) = -\frac{k\alpha^2}{(k+\alpha)^2}(\frac{p}{2} - p)^2 - \frac{k^2\alpha}{(k+\alpha)^2}\int_0^p (p-y)^2 dy/p, \qquad (23)$$

$$U_p(B) = -\frac{k\alpha^2}{(k+\alpha)^2}(\frac{1+p}{2} - p)^2 - \frac{k^2\alpha}{(k+\alpha)^2}\int_p^1 (p-y)^2 dy/(1-p). \qquad (24)$$

Since $p > \frac{1}{2}$, $U_p(B) > U_p(A)$. In addition, because $\bar{y}_B > \bar{y}_A$, $U_z(B) - U_z(A)$ is strictly increasing in $z$. Therefore, there exists $\tilde{p} < p$ such that parents with trait $z \in (\tilde{p}, p)$ in neighborhood $A$ will have an incentive to move to neighborhood $B$ because for them, $U_z(B) > U_z(A)$. This shows that $S_A = [0,p)$, $S_B = [p,1]$ cannot be supported in equilibrium.