Luck and Effort
Learning About Income from Friends and Neighbors*
JOB MARKET PAPER

Gustavo A. Caballero†

University of Calgary

This draft, October 2016. Find the latest version at
http://goo.gl/mlj6YR

Abstract

Can social segregation explain differences in beliefs regarding the role of effort in determining high incomes? I develop a model of lineages of agents learning about the effect of effort on the probability of receiving a high income based on their own experience (effort chosen and income received) and the experiences of agents in their networks. Simulating economies of these agents, only two conditions are needed for the existence of long-run differences in beliefs: (i) agents assume their networks are representative of the whole economy, and (ii) they are more likely to meet others with similar life experiences—as under social segregation. For my analysis, I also consider (iii) imperfect intergenerational transmission of beliefs, in the form of partial confidence about parent's beliefs, to account for sizable changes in beliefs due to single mobility experiences. I find a positive relationship between the degree of social segregation and the level of long-run differences in beliefs. Moreover, high levels of social segregation can lead agents to make inefficient effort choices while average beliefs drift away from real parameters. High levels of social segregation generate groups of agents who persistently exert (no) effort coexisting with agents choosing depending on their cost of effort. And, mobility experiences resulting from luck decrease the belief about the role of effort.

Keywords: Beliefs, social learning, social segregation, political disagreement.

*I am grateful to Robert Oxoby and B. Curtis Eaton for their guidance in this project. Also, I would like to thank Paola Giuliano, Pamela Campa, Lori D. Bougher, Jean-François Wen, participants at the 2016 Canadian Economics Association meetings, the 2016 Midwest Political Science Association conference, the 2014 Behavioral Models of Politics Conference, and seminar participants at the University of Michigan School of Information and the University of Calgary for their valuable comments. Finally, I would like to acknowledge generous financial support by the Vanier Canada Graduate Scholarships provided by the Social Science and Humanities Research Council.

†Ph.D. Candidate, Department of Economics, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada, N2N 1N4. Phone: +1 (403) 220-5857, e-mail: gacaball@ucalgary.ca.

I. Introduction

> One of the things that we [have] learned from common psychology is that preferences and perceptions are more volatile than perhaps classical economists thought. [Therefore, i]t is very useful to go behind them and see if there are more fundamental invariant laws [that] explain what is going on—Mcfadden (2011)

Consider taking a sample of individuals and asking them to what extent they think effort, and not luck, determine incomes. Chances are you will find considerable differences in their answers. In fact, over the course of more than three decades, the World Values Survey (2015) has yet to find a country with a clear majority agreeing on the related question about the role of effort in determining success. Differences in this regard are possible because—as it has been recognized by economists since Vickrey (1945)—income is jointly determined by multiple factors over which there are asymmetries in information. Broadly speaking, income is determined by effort and luck (Alesina and Angeletos, 2005), and luck can introduce a considerable amount of uncertainty into the relationship between effort and income. It is not uncommon to encounter individuals who exert low effort and have high incomes, as well as some exerting high effort and having low incomes.

Individuals, and societies, uncertain about the role of effort in determining incomes make fundamental choices based on beliefs. At the individual level, these choices range from children in the developing world choosing not to go to secondary school because they believe the returns to education are too low (Jensen, 2010), to individuals expressing different attitudes towards redistribution depending on their beliefs regarding the roles of effort and luck in determining incomes (Fong, 2001; Page and Goldstein, 2016). At the aggregate level, differences across countries in average beliefs are significantly correlated with degrees of social spending (Alesina, Glaeser, and Sacerdote, 2001; Alesina and Angeletos, 2005), and "countries where [the] belief [that] effort pays prevails also set harsher punishment for criminals" (Di Tella and Dubra, 2008, p. 1564). Moreover, if beliefs vary so much from one person to the other—as is the case of beliefs regarding the role of effort in determining incomes—considerable differences in beliefs within societies can affect the speed of policy reform (Barber and Mccarty, 2013; Binder, 1999), and even if the policies chosen are optimal (Azzimonti, 2011; Alesina and Tabellini, 1990). Therefore, it is essential to understand the source of individual beliefs and of differences in beliefs among members of a society.

I study whether social segregation can explain the existence of considerable and persistent differences in beliefs regarding the role of effort in determining incomes. When learning about the relationship between effort and income, an individual's natural source of information are the effort choices and incomes received by the people she knows—her network. Yet, under social segregation an individual's network is not representative of the society at large. If the individual fails to account for the bias in the sample of experiences she procures from her network, learning from her network can lead to biased beliefs. I present a model of lineages of agents learning from their parents' beliefs, their own experience (effort exerted and income received), and the experiences of their friends and neighbors, in order to explain the effect of social segregation on the dispersion of beliefs. I argue that persistent differences of beliefs regarding the role of effort in determining incomes result from the interaction between societies' social segregation and individuals' naivete when learning.

In the model, agents (uncertain about the relative roles of effort and luck in determining income) face the decision whether or not to exert costly effort when young to increase the probability of receiving a high income when old. However, low effort does not assure a low income nor does high effort assure a high income. In my model, agents exerting no effort have a positive probability of receiving a high income, and exerting effort increases the probability of receiving a high income. Agents believe effort increases the probability of receiving a high income but are uncertain by how much. They differ in the cost of effort, and the difference in costs is sufficiently high that if agents knew the real probabilistic returns to effort (or held correct beliefs) only agents with low costs find it optimal to exert effort.

Beliefs are central in this model as each generation of agents base their effort choice on the beliefs bequeathed to them by their parents. In the transition to old age, income is realized according to the objective probabilities in the economy and agents update beliefs, before bequeathing beliefs to the next generation. Agents update subjective beliefs received from their parents using their own experience (effort exerted and income received) and information about the experience of agents in their network. Consequently, the model can be understood as one of lineages' learning, where each generation contributes the information they gather when old but agents ignore potential sample biases when updating beliefs. This is important as social segregation biases samples towards each agent's experience. Social segregation is introduced in the model by making the probability of meeting another agent with a certain experience a decreasing function of social distance (as measured both in terms of effort choices and income).

In this environment, beliefs are path dependent and must be traced to all decisions, incomes, and samples that a lineage has observed. This is especially important as some individual beliefs may exceed critical levels after which effort decisions are independent of the possible individual costs. That is, beliefs may evolve in such a way that the choice to exert effort becomes independent of the cost of effort when the returns are excessively over- (under-) estimated. Since the evolution of individual beliefs depends on the choices made by all individuals and the unfolding of risk, finding a general closed form solution for this problem is not possible.[1] Therefore, the analysis presented here is based on two sets of

[1] A solution to the case when effort choices are not affected by beliefs, for example when segregation is very low, is presented in the Appendix B.

simulations of agent economies. Each set of simulations is a collection of economies that differ only by the level of social segregation. The first set of simulations is made under the extreme assumption that all lineages start with the same beliefs, and these initial beliefs are correct. The second set of simulations studies the robustness of the first findings by using techniques developed to study Markov Chain Monte Carlo processes to find the stable state distribution of beliefs.

Simulations show how agents' naivete, when living in segregated societies, leads to significant differences in individual beliefs regarding the returns to effort, even if all lineages start with the same "correct" beliefs. If segregation is very low, there are differences in beliefs but the dispersion of beliefs is small enough that all agents choose effort optimally. However, if segregation is high the dispersion of beliefs is high, extreme beliefs are common and a considerable number of agents make sub-optimal effort choices. In fact, under high levels of segregation agents persistently exerting (no) effort coexist with agents choosing depending on their cost.

As we find beliefs are dispersed in segregated societies, I analyze what experiences lead lineages to different parts of the distribution of beliefs. When segregation is low, and most agents are choosing effort optimally, the main determinant of beliefs are recent draws of costs, whereby lineages drawing low costs more frequently hold the highest estimates on the returns to effort. In contrast, when segregation is high, beliefs are path dependent as extreme beliefs lead some lineages to always or never exert effort, a behavior that reinforce beliefs in this environment. Thus, the experiences lived by generations far in the past of a lineage can have considerable effects on the beliefs and the decisions made by their current members. This is especially important given that in this environment there is no dependence with respect to the costs.

Finally, I study the effects of income mobility in this environment. Piketty (1995) notes that mobility has considerable effects on beliefs. In my simulations, I find that social learning in segregated societies leads poor agents with poor parents to have the lowest estimates on the returns to effort, and rich agents with poor parents to have the highest estimates on the returns to effort. Between these two groups are poor agents with rich parents and rich agents with rich parents, but the order between these two depends on the level of social segregation. In order to understand this pattern, I break down mobility experiences by the effort choices of both agents and their parents, to take into account that luck can also determine a mobility experience. I find that mobility resulting from luck actually decreases beliefs regarding the returns to effort. And, because segregation can change the proportion of agents exerting effort, the average effect of mobility differs between economies due to differences in the frequency of mobility experiences. In all cases, mobility on average does not decrease beliefs as mobility by luck is much less common than mobility resulting from choices.

I focus on social segregation for two reasons. First, when Jensen (2010) discusses why poor children underestimate the returns to education, he notes that "the only data on earnings available to [poor] youths may come from the individuals they can observe around them ...[thus] residential segregation by income could lead to underestimates of the returns to schooling" (p. 516). Second, I provide evidence, at the metropolitan area level in the United States, of a significant positive relationship between the level of income segregation and the level of disagreement regarding the role of effort in determining success. If income segregation leads poor students to underestimate the returns to effort, as suggested by Jensen (2010), income segregation should also lead more affluent students to overestimate the returns to effort. Thus, the difference between the beliefs of these two groups should increase with the level of income segregation, which is what I find for the U.S. when bringing together data on economic segregation from the American Communities Project (2011) and beliefs data from the Trends in Political Values and Core Attitudes 1987-2009 survey by the Pew Research Center for the People & the Press (2009).

To analyze the relationship between the dispersion of beliefs and the level of income segregation, I first calculate the dispersion of beliefs, at the metropolitan area level, as the index of Ordinal Variation of the answers regarding the level of agreement with the statement that "[h]ard work offers little guarantee of success." The index of Ordinal Variation relies on the cumulative distribution to provide a measure of dispersion free of any assumptions about the distribution of the underlying variable determining individual choices. While economists may be less familiar with this measure, it has been widely used by sociologists, who have to deal more frequently with ordinal variables (e.g. Reardon, 2009; Blair and Lacy, 2000). For a measure of income segregation, I rely on the measure provided by The American Communities Project. They measure spatial residential segregation using the rank-order information theory index by Reardon et al. (2006).[2] In short, the rank-order information theory index is proportional to the share of individuals that should be moved between census tracts (an approximation of neighborhoods) in order to make the within-tracts distribution of the rank of incomes resembling the distribution of ranks at the metropolitan area level. By relying on ranks, and not actual incomes, this measure is independent of income inequality.

Figure I presents the relationship between the level of income segregation and the dispersion of beliefs, in U.S. metropolitan areas with at least 5 survey respondents. To take into account for differences in sample sizes (precision), the scatter plot and the linear regression are weighted by the number of survey respondents. There is a significant positive relationship between the level of income segregation and the dispersion of survey responses

[2] A detailed description of these indexes can be found in the Annex A.

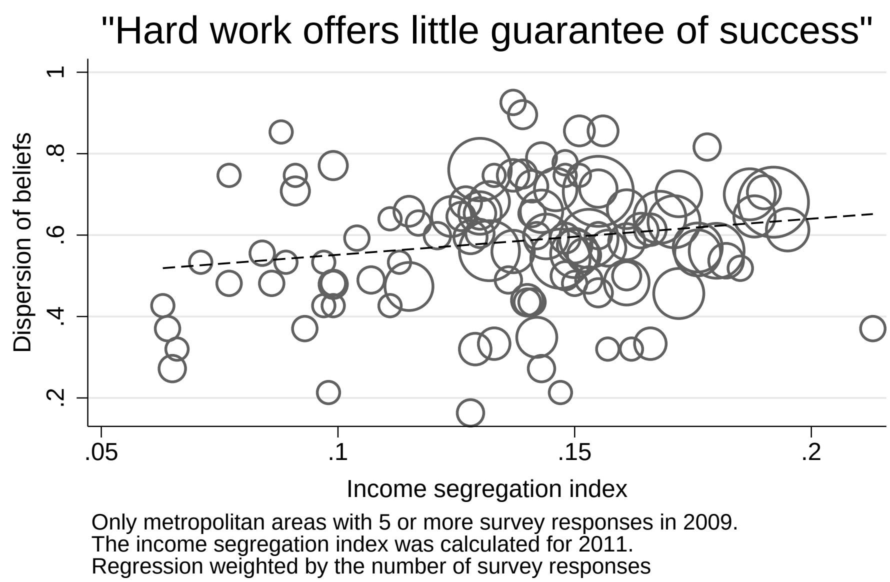


FIGURE I
Relationship between income segregation and the dispersion of beliefs regarding the importance of effort in guaranteeing success (5 or more respondents).

regarding the importance of effort in guaranteeing success. As a matter of fact, this significant positive relationship is found in any given year and with more precisely estimated samples (i.e. restricted to metro areas with at least 10 or 20 survey responses).[3] Moreover, there is a significant and positive relationship between income segregation and the dispersion of beliefs after considering differences in sample sizes (by bootstrapping samples of the same size), metropolitan areas random effects, and in some instances the alternative explanation of income inequality. All these findings are presented in the Annex A.

It is noteworthy that models explaining dispersion of beliefs within societies, beyond selfish motives, are scarce. Much has been written about the source of differences between groups (see Alesina and Giuliano, 2010, and the references therein), but not why beliefs vary within groups. In fact, most of the literature refers to the model developed by Piketty (1995) when discussing the source of disagreements within societies. In his model, individuals learn about the relative roles of effort and predetermined factors only from the past history of their lineage. Similar individuals coming from lineages with different mobility stories hold different beliefs regarding the role of individual effort in determining income.

[3]Considering only the metro areas with at least 5 survey responses, there is information of income segregation and survey responses for 145 metropolitan areas in 1990, 107 in 2000, and 107 in 2010. These regressions are not presented here, but are available on demand.

For example, individuals from a lineage moving down in the income distribution will reveal a lower preference for redistribution when compared with others consistently poor. However, the model's predictions relies on the assumption that individuals do not learn from the experiences of others. This assumption is reasonable, as it is impossible to learn about the experiences lived by everyone in an economy. However, it is possible to learn about the experiences of those in our network—our friends and neighbors. Unlike Piketty (1995), allowing for social learning, the model developed here generates significant levels of disagreement within societies.

Following this introduction, Section II presents a brief overview of the social learning literature and argues for the use of observational social learning in our model. Section III lays out the model. Section IV presents the findings of simulating economies of agents behaving as in the model under varying degrees of segregation. Section V concludes.

II. Learning about income from others

As noted in the introduction, segregation can affect beliefs as it shaped the set of individuals one has access to when trying to resolve an uncertainty. Individuals uncertain about the effect of effort on their chances of receiving a high income can use information obtained from others, in addition to their own experience, to decrease such uncertainty. Using information gathered from others when learning is what economists denote as Social Learning (Mobius and Rosenblat, 2014).

To learn about the relationship between effort and income, individuals can engage in different types of social learning. For example, they may share their priors to update beliefs as in DeGroot (1974), where consensus over subjective distributions of a parameter can be achieved by iteratively updating beliefs using an individual's prior and information about the beliefs of others, even if more weight is given to those with more experience or knowledge about the subject. Learning in this fashion, and under very few conditions, individuals are expected to converge towards a consensus. A notable exception is Arifovic, Eaton, and Walker (2015) who demonstrate that, by sharing priors, individuals can significantly and persistently disagree if individuals choose neighbors in order to minimize cognitive dissonance.

Learning by sharing priors is arguably preferred only when experimentation is not possible. In other words, one can expect individuals relying more on experience than opinions, if they can experiment. In an overlapping generations model, individuals can learn from their parents' experience. Moreover, individuals can also learn from the experiences of others observed by their parents.

The current model is one of observational learning. Observational learning has the main advantage that it provides individuals with information beyond their own experience every period. Moreover, learning by observing others provides information about possible outcomes when making choices believed to be sub-optimal. As Piketty (1995) highlights, “completely learning the relative role of effort in the generation of inequality would require a lot of costly experimentation that each single generation is not willing to undertake.” By engaging in social learning, lineages do not need to undertake all the experimentation themselves as they can obtain this information by observing others. Moreover, learning from the experiences of others allow for “faster” learning.[4]

It is noteworthy that, learning from others’ choices and outcomes, albeit frequently observed, is an understudied type of social learning.[5] This is partly because social learning of this type is believed to almost certainly lead to consistent estimates of underlying parameters (Gale and Kariv, 2003), even if individuals are not sophisticated learners (Bala and Goyal, 1998). Moreover, even if the unknown parameters are time varying and individuals are not sophisticated, allowing subjects to observe the full experience of others is believed to lead to convergence (in an statistical sense) on accurate beliefs (Acemoglu, Nedic, and Ozdaglar, 2008). Therefore, this paper also contributes to the social learning literature by presenting a situation when observational social learning can lead to inaccurate beliefs.[6]

Learning about others’ effort choices and income (experiences) is possible through social interactions, and individuals use observations from their networks to inform their beliefs. If individuals are able to observe a representative sample of society through their interactions (or they are able to recognize any biases their samples have), by using the information of others’ experiences individuals should be able to reach consistent estimates of the returns for their effort. Consequently, two fundamental questions here are (i) how individuals’ samples are procured, and (ii) how do individuals use their samples to make inferences about the relative roles of luck and effort in determining income.

[4] Updating beliefs as proposed by the model implies that, the uncertainty regarding underlying parameters, i.e. the variance of individual beliefs themselves, is inversely related to the number of observations one can obtain. Therefore, by engaging in observational learning agents will decrease their variance (increase their confidence) more rapidly.

[5] An intermediate case, also extensively studied in the literature, is one named by Mobius and Rosenblat (2014) as observational learning, that I would call actions observational learning. There is only one state of the world and individuals can only observe the actions made by others and make a once-and-only action. This literature has been instrumental to study “herd” behaviour (e.g. Banerjee, 1992).

[6] Without the capacity to present a general proof, my conjecture is that the unbiasedness of long-run beliefs found in previous papers result from the common assumption that signals, while noisy, are on average unbiased. Under segregation, even though individual experiences are on average unbiased, the signals individuals receive from their networks are not.

III. A model of naive social learning in segregated societies

III. A Rationale

Individuals' friends and neighbors are not usually representative of the population. Rather, individuals tend to be spatially close, in their residence and consumption patterns, to those with similar "life experiences" (Schelling, 1971). This results from individuals' preference to live near others with similar innate characteristics, who have made similar choices, and who have obtained similar incomes.[7] In fact, social segregation characterizes current societies. For example, economists and sociologists alike have noted that the U.S. is highly segregated by multiple factors, including education levels and income (Cutler and Glaeser, 1997; Massey, Rothwell, and Domina, 2009).[8] What is more, in the U.S., income segregation has been increasing in the recent decades (Reardon and Bischoff, 2011; Bischoff and Reardon, 2014). This is important as individuals meet more frequently with their neighbors who, in segregated societies, are closest in social status, education and income (Akerlof, 1997; Scheinkman, 2008). Therefore, when living in socially segregated societies, individuals' networks are over-representative of their own life experience. In my model, individuals learn based on the information gathered from segregated networks.

When learning about the determinants of income, non-representative networks can be problematic as individuals have a tendency to attribute representativeness to their own observation of uncertain phenomena (Tversky and Kahneman, 1971, 1974). As Kahneman (2013) puts it, "[individuals have a] strong bias toward believing that small samples closely resemble the population from which they are drawn." As a result, individuals are inclined to be naive and ignore sample biases resulting from small samples.[9] Tversky and Kahneman (1974) refer to this behaviour as believing in "The Law of Small Numbers" (LSN, hereafter) capturing the use of a heuristic—an intuitive rule of behavior minimizing cognitive effort by assuming small samples are representative of the population. Indeed, Kahneman (2013) concludes that "[w]e easily think associatively, we think metaphorically, we think causally, but statistics requires thinking about many things at once, which is something that [our intuitive] system is not designed to do" (p. 13).[10]

[7] A slight preference—which could even be subconscious—to live close to similar peers is sufficient to generate a significant degree of segregation.

[8] Race is still the primary source of segregation in the U.S. However, ever since the 1950's, segregation by race has been declining to be slowly replaced by income-based segregation. Iceland and Wilkes (2006) argue that the current level of racial segregation is mainly a result of the correlation between race and income. Specifically, they find that "in both 1990 and 2000 high-[Socio-Economic Status] racial and ethnic groups were significantly less segregated from non-Hispanic whites than corresponding low-SES groups."

[9] While a reader may consider that this bias is only observed on the statistically unsophisticated, it is noteworthy that Tversky and Kahneman (1974) study the responses of professional psychologists.

[10] In the bounded rationality literature, cognition is divided into two systems. First there is an intuitive system with spontaneously coming thoughts made with minimal computation, effort or consciousnesses.

Notice that neither naivete nor social segregation are problematic on their own. On the one hand, if individuals are able to obtain truly random samples through their social interactions, then the heuristic and reality coincide. In fact, obtaining random samples on average would be sufficient for consistency in the current model. Moreover, by using a heuristic, instead of analyzing the sample problem, individuals would minimize cognitive effort while still reaching consistent estimates. On the other hand, procuring biased samples through social interactions might not be a problem if individuals are aware of the bias. That is, one can suppose that—similar to a social scientist—individuals may be able to correct for biases in their samples if they are aware of the non-representativeness of their networks. Hence, it is a combination of individual specific sample biases and the application of the LSN that allows for significant differences in beliefs.

While I find that individual's naivete when learning and sample biases, arising from social segregation, are sufficient conditions to explain long-run disagreements, I also consider imperfect intergenerational transmission of beliefs to account for the central role of mobility experiences in shaping individuals' beliefs. If beliefs are perfectly transmitted between generations, the process of learning implies that in the long-run experiences of any single generation should not have a distinguishable effect on beliefs. However, in the case of beliefs about the role of effort in determining incomes, there is strong evidence that individual beliefs are related to single experiences of income mobility (Piketty, 1995). This can be explained by the fact, while parents' beliefs and attitudes are good predictors of individuals' beliefs and attitudes, individuals' beliefs and attitudes do not perfectly reflect those of their parents.[11] Among the many reasons that have been proposed to explain why beliefs transmission is imperfect, in the model I consider a way of children's skepticism (Mills, 2013; Harris and Corriveau, 2011) by making agents less confident than their parents about the beliefs they receive intergenerationally.

III. B Model Description

In this section, I develop an understanding of how segregation and individual naivete can affect individuals' learning about the role of effort in determining income. I present a model of overlapping generations, with no population growth, where agents from every generation are confronted with the choice to exert (or not) costly effort when they are young.

Exerting effort increases the probability of receiving a high income when an agent is old. Income is only high or low, and the difference between high and low incomes, weighted by

Second, there is reasoning system which requires a significant amount of cognitive effort (Kahneman, 2003). Representativeness heuristics are part of the intuitive system.

[11] For example, Dohmen et al. (2012) provide evidence of a strong and positive correlation between children and parent's risk and trust attitudes. For references of more evidence regarding imperfect intergenerational transmission of beliefs see Bisin and Verdier (2005)

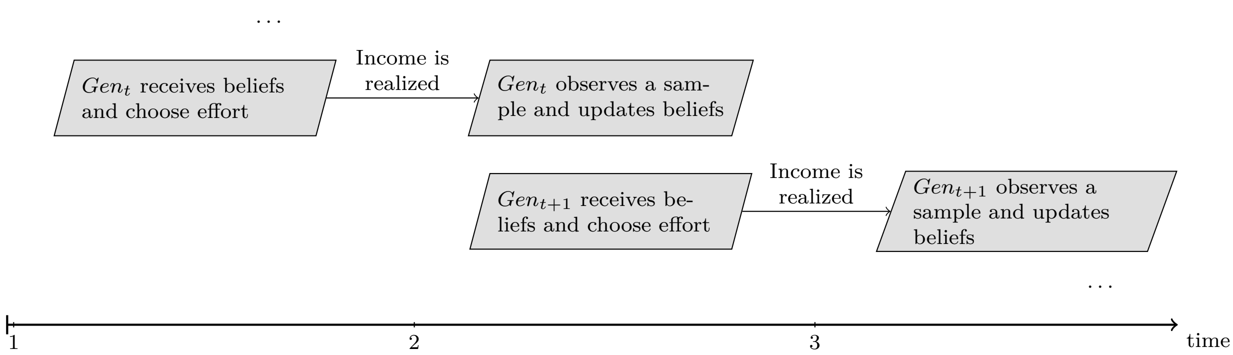


FIGURE II
Model timing.

the increase in probability associated to effort, corresponds to the expected returns to effort. All agents face the same probability of receiving a high income when exerting no effort and a unique higher probability when exerting effort. Hence, the probabilities conditional on effort are the same for all individuals in the economy. However, the probabilities of receiving a high income associated with each level of effort are never one or zero. Namely, some of the agents exerting no effort may be "lucky" and receive a high income, while others exerting effort may be "unlucky" and receive a low income. The presence or absence of effort does not guarantee a high or low income.

My model differs from others in which groups of individuals have unequal returns on effort or chances of upward luck (e.g. Piketty, 1995). This is not to say that differences in opportunities between groups or the actual returns to effort are not fundamental in explaining diversity of beliefs in societies. Differences in opportunities or the returns to effort potentially affect individual beliefs regarding the role of effort and luck in determining incomes. The main contribution of this model is to present an explanation of why beliefs are expected to vary even in relatively homogeneous societies.

The model is dynamic in that beliefs are formed at the lineage level by a succession of generations making effort choices and learning about the processes determining incomes. When agents are young, they are bequeathed a set of beliefs by the old of their lineage and use these beliefs to make their effort choice. In their transition to old age, income is realized according to the objective probabilities of the economy and the effort exerted. The effort chosen and income received determines an agent's experience, and agents learn about other experiences through social interaction. However, the network that an agent meets—their sample—over-represents the agent's experience due to social segregation. Having procured a sample through social interaction, agents update beliefs in a Bayesian manner and bequeath their updated beliefs to their descendants. When updating beliefs, agents are

TABLE I

DETERMINATION OF INCOME

		Prob. of ω_h	Prob. of ω_l
Effort	1	$\theta+\delta$	$1-(\theta+\delta)$
	0	θ	$1-\theta$

not fully Bayesian as they do not consider the sample bias resulting from segregation.

The timing of decisions, procurement of samples, and updating of beliefs is presented in Figure II for generations t and $t+1$. Below, I provide a detailed description of the model.

The economy

Agents live in an environment where (i) income can only be high or low, that is $\omega_i^t \in \{\omega_h, \omega_l\}$, (ii) there is always a chance $\theta \in (0,1)$ to earn a high income even if effort is not exerted, (iii) individual chances increase by $\delta \in (0, 1-\theta)$ if costly effort is exerted, and (iv) even if effort is exerted there is a chance of $1-\theta-\delta$ of earning a low wage. Table I summarizes this economy. In this model, both the true θ and δ are unknown so agents must use their beliefs (i.e. estimates of $\hat{\theta}_i^t$ and $\hat{\delta}_i^t$) to make the effort decision every period. While exerting effort increases the chance of receiving a higher income (i.e. $\delta > 0$), effort has a cost. This means that not knowing the actual returns to effort can be costly as (i) agents could fail to exert effort when the cost is actually lower than the expected returns to effort, or (ii) they could be exerting effort when the expected returns are less than the cost of effort.

Agents' effort choices

The problem to be solved by agents of generation t from lineage i is to choose whether to exert costly effort $e_i^t \in \{0,1\}$ when young in order to maximize lifetime consumption. This decision is made conditional on a randomly determined individual cost of effort $\xi_i^t \in \{\xi_l, \xi_h\}$ which is only privately known. I assume costs are randomly determined every generation and draws are independent between generations. The difference in costs is such that, if agents knew the real δ, only those with a low cost would exert effort. Additionally, I assume that the probability of drawing a low cost (ξ_l) is not related in any way to the lineage. Namely, all agents from each generation face the same probability λ of obtaining ξ_l.

Agents' consumption when old, $c_{i,t+1}^t$, is an exogeneously determined portion $(1-\eta) \in [0,1)$ of their earnings ω_i^t, leaving the remaining portion η for the consumption of their descendants.[12] That η is exogeneously determined implies there is no intertemporal choice

[12]This implies that they consume the same portion of their parents' income when young.

in consumption and generations do not interact actively beyond the transmission of beliefs. However, the model includes this income sharing rule within lineages to provide a source of income to the young. This income sharing rule makes the model flexible enough to allow, for example, decisions to invest in modifying the probabilities faced by subsequent generations. Similarly, I assume there is no savings and no time discounting.

To summarize, the problem to be solved when agent i is young is:[13]

$$\begin{aligned} \max_{e_i^t \in \{0,1\}} u_i^t =& c_{i,t}^t + c_{i,t+1}^t - \xi_i^t e_i^t \qquad \text{s.t.} \\ c_{i,t}^t \leq& \eta \omega_i^{t-1} \\ c_{i,t+1}^t \leq& \begin{cases} (1-\eta)\omega_l \text{ with prob.} = 1 - \hat{\theta}_i^t - \hat{\delta}_i^t e_i^t \\ (1-\eta)\omega_h \text{ with prob.} = \hat{\theta}_i^t + \hat{\delta}_i^t e_i^t \end{cases} \end{aligned} \tag{1}$$

In particular, agents use beliefs regarding θ and δ when solving their problem as these are their best estimates of the parameters. Given all constraints bind, this problem can be restated as

$$\max_{e_i^t \in \{0,1\}} E[u_i^t] = \eta \omega_i^{t-1} + (1-\eta)\left(\omega_l + (\hat{\theta}_i^t + \hat{\delta}_i^t e_i^t)(\omega_h - \omega_l)\right) - \xi_i^t e_i^t \tag{2}$$

Consequently, if I assume that agents decide to exert effort when indifferent, an agent exerts effort when young (i.e. $e_i^t = 1$) if

$$\hat{\delta}_i^t (1-\eta)(\omega_h - \omega_l) \geq \xi_i^t. \tag{3}$$

That is, agents exert effort if the increase in earnings, weighted by their belief regarding δ, is greater than the individual cost drawn by generation t. Notice that $\hat{\theta}_i^t$ does not matter for the decision as their probability to receive a high income for both levels of effort include this factor. As stated before, this decision rule implies that if an agent has a sufficiently high belief regarding δ then she exerts effort even if the cost is high. Similarly, an agent with low enough beliefs exerts no effort, even if she faces a low cost of effort. Throughout the paper I refer to such cases as "extreme beliefs."

[13] Because the cost of effort appears in the utility function and not in the budget constraint, all individuals who desire to assume the cost can do it, therefore the decision to save has no significant implications. Rabin (1995) discusses the difference between having the cost as a constraint or in the utility function. Galor and Zeira (1993) study the case when the cost of effort is financial and agents are financially constrained in an education decision.

Network formation and segregation

When agents are old, they form a network that serves as a sample of effort-income combinations (experiences), in addition to their own experience. Then, they use this sample of experiences to update beliefs. In this model, there are only four possible experiences: (i) having exerted effort and receiving a high income, (ii) exerting effort and receiving a low income, (iii) not exerting effort and receiving a high income, and (iv) exerting low effort and receiving a low income. By living in a segregated economy, the sample each agent observes is biased towards their own experience.

Segregation in the model is captured by assuming agents interact the most with others who have the same level of effort and receive the same income, and interact the least with people that chose different effort and receive a different income.[14] Whether agents interact more with agents exerting the same effort, but receiving a different income, or with those who exert a different level of effort, but receive the same income, depends on the type of society modeled. I model segregation by setting the probability that an individual exerting effort e_i^t and receiving income ω_i^t meets another person exerting effort e_j^t and receiving income ω_j^t as:

$$\psi_{i,j}^t = \psi^{\gamma_e |e_i^t - e_j^t| + \gamma_\omega |I^h[\omega_i^t] - I^h[\omega_j^t]|} \tag{4}$$

where $\psi \in (0, 1)$ and the parameters $\gamma_e \geq 0$, and $\gamma_\omega \geq 0$ capture how much segregation there is with respect to effort and income. $I^h[\cdot]$ is an indicator of whether the agent received a high income or not. Therefore, the more socially distant two agents are, the less likely they meet. This way of modeling segregation is equivalent to assigning a probability based on each characteristic, and assuming that these probabilities are independent. Yet, the current approach allows for a single definition of segregation, and is thus more appropriate for our analysis.

To observe the effects of segregation in the procurement of samples, let's consider an economy of 512 agents distributed, in terms of the experiences lived, as shown in Table II. The table shows that, for example, 256 agents exerted effort and received a high income. Assume segregation is such that $\psi = 1/2$, $\gamma_e = 1$ and $\gamma_\omega = 2$. Hence, agents with different stories will observe samples as presented in Table III. For example, an agent exerting effort and receiving a high income would observe all other 255 agents with her same experience (in addition to herself), 16 other agents having exerted effort but receiving a low income, 32 agents who exerted no effort but received a high income, and 16 agents exerting no effort and receiving a low income.

[14] This network can be understood as the equilibrium network of costly communication between agents as in Jackson and Wolinsky (1996).

TABLE II

EXAMPLE: ACTUAL DISTRIBUTION OF TYPES

		Income ω_h	Income ω_l
Effort	1	256	64
	0	64	128

TABLE III

EXAMPLE: SAMPLES PROCURED BY INDIVIDUALS ACCORDING TO THEIR OWN EXPERIENCE WHEN $\psi = 1/2$, $\gamma_e = 1$ AND $\gamma_\omega = 2$

(a) $(e_i^t, w_i^t) = (1, \omega_h)$

		Income ω_h	Income ω_l
Effort	1	256	16
	0	32	16

(b) $(e_i^t, w_i^t) = (1, \omega_l)$

		Income ω_h	Income ω_l
Effort	1	64	64
	0	8	64

(c) $(e_i^t, w_i^t) = (0, \omega_h)$

		Income ω_h	Income ω_l
Effort	1	128	8
	0	64	32

(d) $(w_i^t, e_i^t) = (\omega_l, 0)$

		Income ω_h	Income ω_l
Effort	1	32	32
	0	16	128

Updating Beliefs

An agent from generation t of lineage i holds a system of beliefs, which she uses to make her effort choice. At the core of her system of beliefs are two objects. First, there is the belief regarding the probability of receiving a high income when exerting no effort (i.e. $\hat{P}_i^t(w_i^t = w_h | e_i^t = 0)$) denoted by $\hat{\theta}_i^t$. Second, there is the probability of earning a high income when exerting effort (i.e. $\hat{P}_i^t(w_i^t = w_h | e_i^t = 1)$) corresponding to $\hat{\theta}_i^t + \hat{\delta}_i^t$. Interestingly, an agent exerting effort may receive a high income because she was going to be lucky in any case or because she exerted effort. Based on these two beliefs, one can solve for other parameters of interest. Specifically, one can solve for the probability of receiving a low income when exerting effort, which is $1 - (\hat{\theta}_i^t + \hat{\delta}_i^t)$, and more importantly the increase in the probability of receiving a high income associated with exerting effort

$$\hat{\delta}_i^t = \hat{P}_i^t(w_i^t = w_h | e_i^t = 1) - \hat{P}_i^t(w_i^t = w_h | e_i^t = 0). \tag{5}$$

The system of beliefs is updated every period in a Bayesian manner based on the sample of experiences procured. However, agents are bounded in their rationality and assume their samples are random. Bayesian updating states that an agent's updated beliefs (i.e. the posterior distributions $p_i^t(\theta|y)$) are proportional to her prior belief $p_i^{t-1}(y|\theta)$ and the

TABLE IV
BAYESIAN UPDATING OF A BETA DISTRIBUTION WHEN THE LIKELIHOOD IS DISTRIBUTED AS A $B(n, p)$

Statistic	Prior	Posterior
Mean	$\frac{a}{(a+b)}$	$\frac{(a+pn)}{(a+b+n)}$
Mode	$\frac{(a-1)}{(a+b-2)}$	$\frac{(a+pn-1)}{(a+b+n-2)}$
Variance	$\frac{ab}{(a+b)^2(a+b+1)}$	$\frac{(a+pn)(b+(1-p)n)}{(a+b+n)^2(a+b+n+1)}$

likelihood she observes given the sample obtained that period $p_i^t(\theta)$.

$$p_i^t(\theta|y) \propto p_i^{t-1}(y|\theta) \times p_i^t(\theta) \tag{6}$$

Uncertain about θ and δ, agents do not hold a unique estimation, but instead their beliefs are captured by a probability distribution. For this purpose, I assume that an agent's beliefs regarding these parameters can be captured by Beta distributions $\boldsymbol{\beta}$ (a, b), commonly used in models of Bayesian learning when the parameter to be estimated is a probability. The left column in Table IV presents relevant statistics of a Beta distribution. Two moments of special interest are the mean and the mode. They give the decision maker a unique estimation which can be used when making decisions.[15] Note, the variance is decreasing in the parameters a and b, and hence agents observing larger samples will have a lower variance in their estimates.

In this case, the likelihood is based on the proportion of agents observed earning a high income for each level of effort. Given that there are only two possible incomes, the likelihood of receiving a high income (conditional on a level of effort) can be described by a binomial distribution $B(n, p)$ with parameters n (the number of individuals observed in the given level of effort) and p (the proportion of individuals earning the high income). This means that an agent's observation of high and low incomes, for a given level of effort, resembles an experiment in which successes and errors can separated.

An advantage of using a Beta distribution as a prior is that a Beta distribution is

[15]Which to choose when deciding depends on a behavioural assumption about agents process of recalling beliefs. Either the agent acts based on the estimate which more frequently comes into her mind, the mode, or she is able to find the mean. For big numbers of a and b any difference is negligible. In the simulations I use the mean.

conjugate to a binomial likelihood, so the distribution of the posterior is of the same family as the prior. Bayesian updating in this model can be described by:

$$\boldsymbol{\beta}_i^t\,(a + pn, b + (1-p)n) \propto B_i^t(n,p) \times \boldsymbol{\beta}_i^{t-1}(a,b) \tag{7}$$

That is, for every level of effort the number of agents observed in one's sample who receive a high income are "added" to the a parameter, while those receiving a low income are added to the b parameter. The right column in Table IV presents the mean, median and variance of the updated belief. Using this procedure, individuals update their beliefs for the two possible levels of effort.[16] It is noteworthy that segregation provides samples over-representing the agent's experience. For example, it implies that an agent from a lineage with a long history of exerting effort and obtaining high incomes will be less confident about her estimates of the chance that someone exerting low effort has to obtain a high income.

Finally, young's skepticism is introduced by considering a belief's confidence factor $\rho \in [0,1]$ so beliefs are updated as follows:

$$\boldsymbol{\beta}_i^t\,(\rho a + pn, \rho b + (1-p)n) \propto B_i^t(n,p) \times \boldsymbol{\beta}_i^{t-1}(\rho a, \rho b), \tag{8}$$

Going back to the situation presented in Table III, consider an agent, who exerted effort and received a high income, updating her belief regarding $\theta + \delta$. Further assume that the objective probability is 70%, and that the agent holds a "correct", but weak, prior (i.e. $a_{t-1} = 7$ and $b_{t-1} = 3$). Finally assume that the agent discounts her parent beliefs with a $\rho = 0.5$. In this case, she would update beliefs as:

$$\boldsymbol{\beta}_i^t(259.5, 17.5) \propto B_i^t(272, .94) \times \boldsymbol{\beta}_i^{t-1}(0.5 * 7, 0.5 * 3), \tag{9}$$

resulting in an updated mean belief $\hat{\theta} + \hat{\delta} = 0.935$.

IV. Simulations

This section presents the results of two different sets of simulations of agent economies.[17] In the first set of simulations, I take the extreme assumption that all agents of the first

[16] Assuming this type of learning implies that, waiting until the sample is complete before updating beliefs is equivalent to updating after every interaction. In fact, updating after every interaction can be thought as if, for every experience an agent observes the agent slightly—the size of the effect actually would depend on how "confident" one is about its prior—increases or reduces the belief regarding the probability of earning a high income for that level of effort. Nonetheless, it can be argued that recognizing sample biases when updating after every interaction is highly unlikely.

[17] An analytic solution for the case when effort choices are independent of beliefs and beliefs are perfectly transmitted between generations is presented in the Annex B.

generation hold correct beliefs, $\hat{\delta}_i^0 = \delta \quad \forall i$, but with a minimal level of confidence.[18] In particular, I study the distribution of beliefs after $g = 1,000$ generations of agents, with a particular focus on the last 100 generations. In the second set of simulations, I use the method by Gelman and Shirley (2011) to find the stable state for each level of segregation, and study the distribution of beliefs under the stable state. Within each set, simulated economies only differ by the level of social segregation. In all cases, I simulate the evolution of beliefs of $n = 500$ lineages living in the following environment:

- Possible incomes: $\omega_l = 1$; $\omega_h = 5$.
- Probabilities of receiving a high income: $\theta = 0.2 > 0$; $\theta + \delta = 0.7 < 1$, so $\delta = 0.5$.
- Costs: $\xi_l = 1.5$; $\xi_h = 2.5$; The probability of receiving a low cost $\lambda = 2/3$ (and i.i.d.).
- Segregation by effort $\gamma_e = 1$ is lower than segregation by income $\gamma_\omega = 2$.
- All income is consumed by the parents, $\eta = 0$.
- Agents are confident on the beliefs they receive by a factor ρ of 0.5.[19]

Notice, while this is not intended to be a calibration exercise and the model refers generally to effort choices, one particular effort choice of interest is the decision to continue studying beyond high school. The 2009 High School Longitudinal Study (HSLS:2009, hereafter) by the National Center for Education Statistics (2016) provides useful information in this regard. According to the HSLS:2009, 30% of households with a household head holding a high-school degree or less have a household income above the median (\$50.000), while the percentage is 75% in the case when the household head has more than high-school. Also, about 60% of the students in the HSLS:2009 go on to college. Similarly, how income and effort affect the probability of meeting an agent of a different type reflects the fact that the U.S. is more segregated by income than by education levels (Massey et al., 2009). Therefore the assumptions used in these simulations—though admittedly arbitrary—are aligned with some basic characteristics of the U.S.

I consider four levels of overall segregation $\psi \in \{.2, .4, .6, .8\}$ which I refer to as very high, high, low and very low levels of segregation. In the simulations I apply the Law of

[18] Assigning an uninformative prior to the first generation does not have any significant impact on the behavior here described as a weak prior (not very confident) is quickly overridden by observed data.

[19] While the model has been constructed in terms of transmission of beliefs, it can also be understood as one where agents hear the stories directly from their parents, and their parents' parents, etc. Yet the probability one hears about a particular experience—directly from the predecessor that observed it, or indirectly by another predecessor—falls very quickly. A confidence factor ρ of 0.5 is such that beliefs are mostly build upon the information from three to four generations, which is consistent with both alternatives. Results with perfect intergenerational transmission of beliefs are presented in Annex C.

Large Numbers, so agents meet a portion of agents with a certain experience equal to the probability of meeting an agent of that type.

IV. A Beliefs of lineages starting with the same "correct" prior.

I begin by analyzing the evolution of the distribution of core beliefs, and the distribution of the implied $\hat{\delta}_i^t$ when all lineages start with the same "correct" prior. Box plots are used to summarize the distribution of beliefs in every period. These plots present a box bounded by the 25th and 75th percentiles of the distribution. Additionally, the median is highlighted, and whiskers are extended to the closest of (i) the maximum and minimum or (ii) 1.5 times the distance between the 25th and 75th percentiles. Any observation beyond the whisker has its own marker.

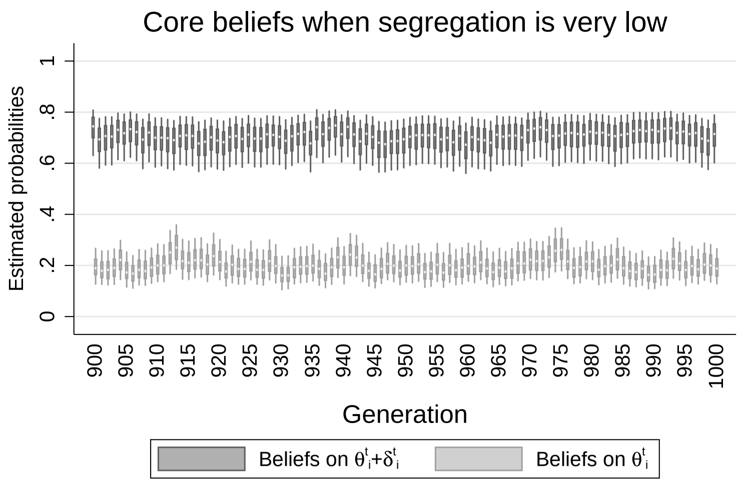
Figure III presents how agent's beliefs regarding θ and $\theta + \delta$ evolve under different levels of segregation. It is striking to observe how, even low segregation is enough to start observing significantly different beliefs. Living in segregated societies leads to significant differences in beliefs regarding θ and $\theta+\delta$, and the variance of beliefs (level of disagreement) is monotonically increasing with the level of social segregation.

How do differences in core beliefs translate to differences with respect to the returns to effort? To answer this question we can use Equation (5) and solve for $\hat{\delta}_i^t$, the variable affecting effort decisions. Figure IV shows how beliefs regarding δ evolve as well as the proportion of agents exerting effort. Remember that if agents were able to know the actual δ, two out of every three should be exerting effort. I highlight an area of efficient choices, the range of beliefs where agents choose effort optimally. Outside this area, agents are excessively over- (under-) estimating the returns to effort so they exert (no) effort under any of the possible cost conditions; namely, agents whose beliefs are outside this area are holding extreme beliefs and are likely to make inefficient effort choices. With the parameters of the simulation and using Equation (3), agents holding a belief $\hat{\delta}_i^t \geq 0.625$ will always exert effort, and will never exert effort if $\hat{\delta}_i^t \leq 0.375$.

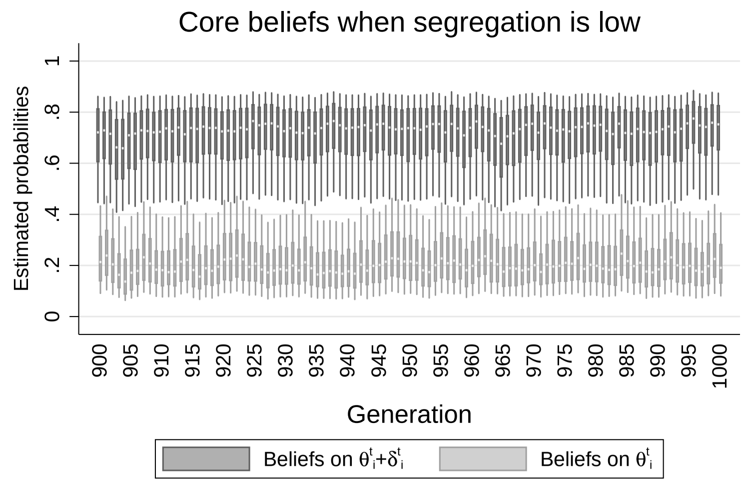
While extreme beliefs regarding δ are seldomly observed in low levels of segregation (i.e. $\psi = .6$), as shown in Figure IVb, when segregation is high or very high (when ψ is low) extreme beliefs are common and affect a significant number of decisions to exert effort. As a consequence, when segregation is high or very high the agents in the economy are on average exerting sub-optimal levels of effort due to extreme beliefs.

IV. B Beliefs under the stable state

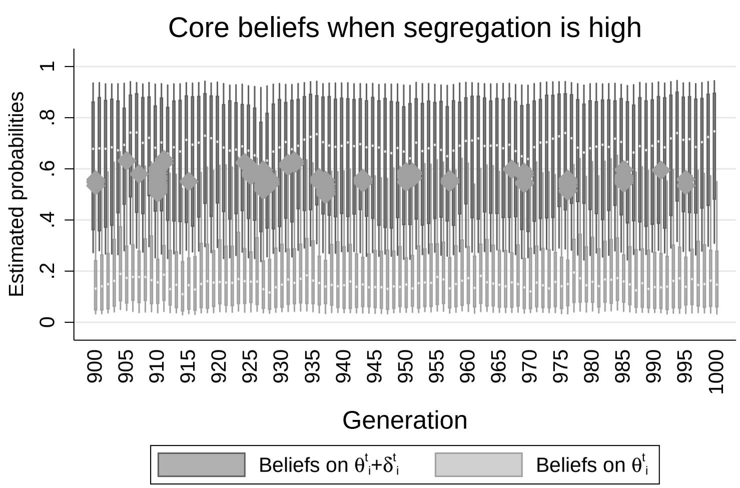
Here, I relax the assumption that all lineages start with the same beliefs, to find a stable state under these conditions, and the characteristics of beliefs and behavior under the stable



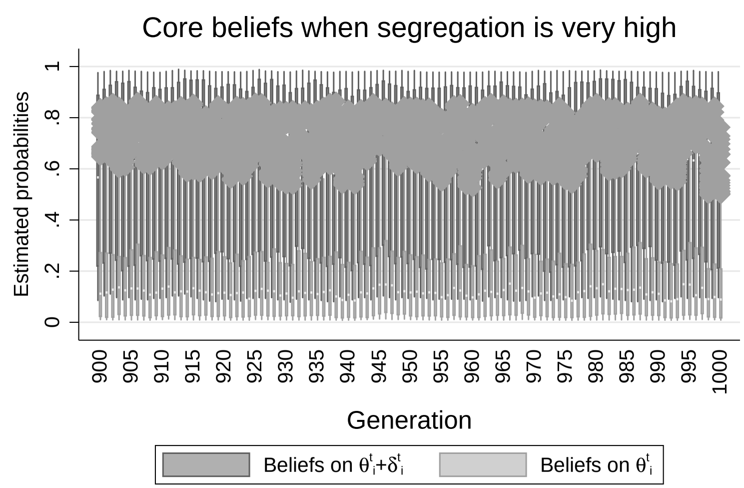
(A) $\psi = .8$



(B) $\psi = .6$

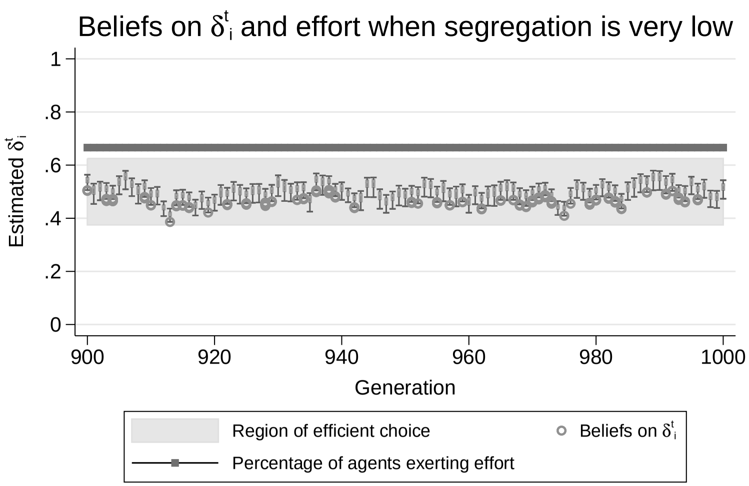


(C) $\psi = .4$

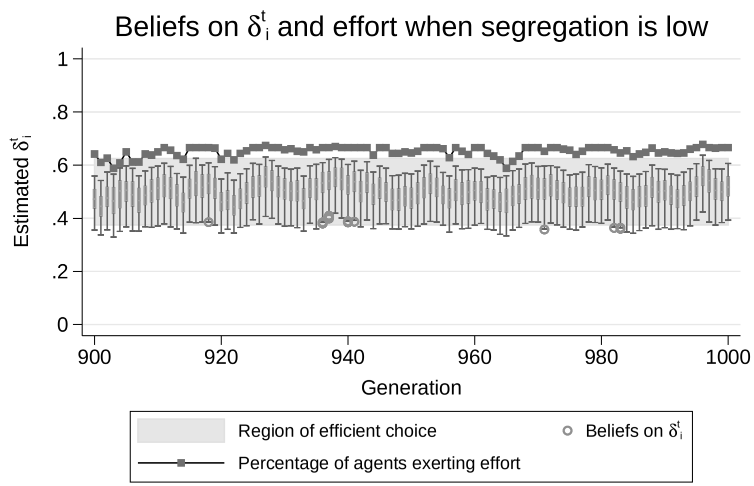


(D) $\psi = .2$

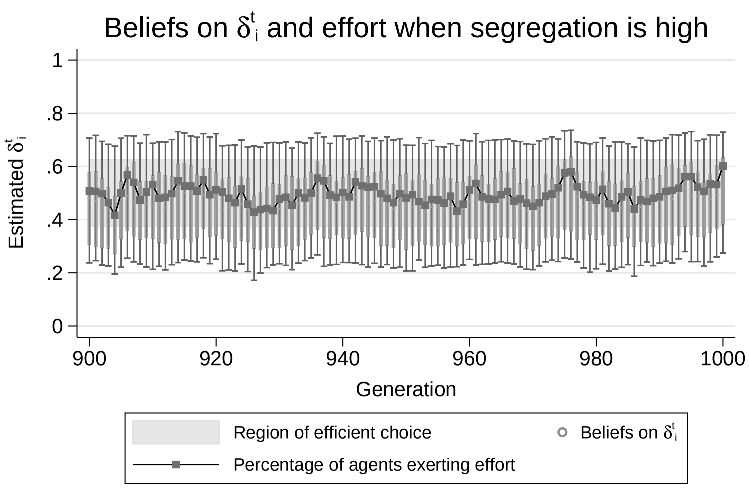
FIGURE III
Evolution of beliefs regarding θ and $\theta + \delta$.



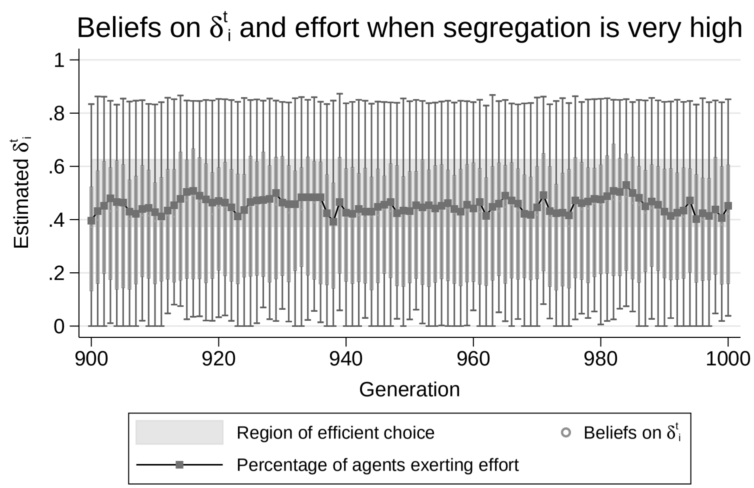
(A) $\psi = .8$



(B) $\psi = .6$



(C) $\psi = .4$



(D) $\psi = .2$

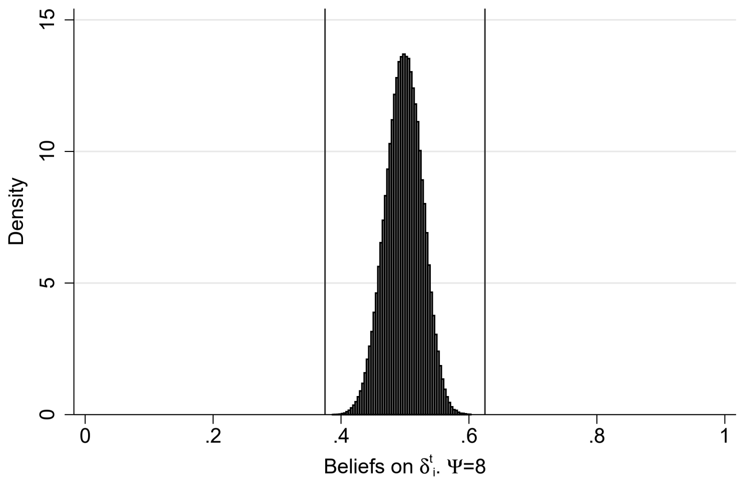
FIGURE IV
Evolution of beliefs regarding δ.

state. A stable state is defined as a bounded distribution of beliefs with stable variance and coefficient of variation (CV). Thus, beliefs at the agent level might change but on average they remain the same. I follow the methodology described by Gelman and Shirley (2011) to study Markov Chain Monte Carlo (MCMC) processes. It is reasonable to use techniques developed for MCMC as the learning model here developed is essentially a MCMC. That is, beliefs in period $t+1$ depend only on the beliefs in period t and a random error. This is due to the fact that beliefs in period t summarize all the information gathered in previous periods. Therefore, one can expect that the conditional distribution of beliefs in period $t+1$, given the beliefs in period t, does not depend on t.

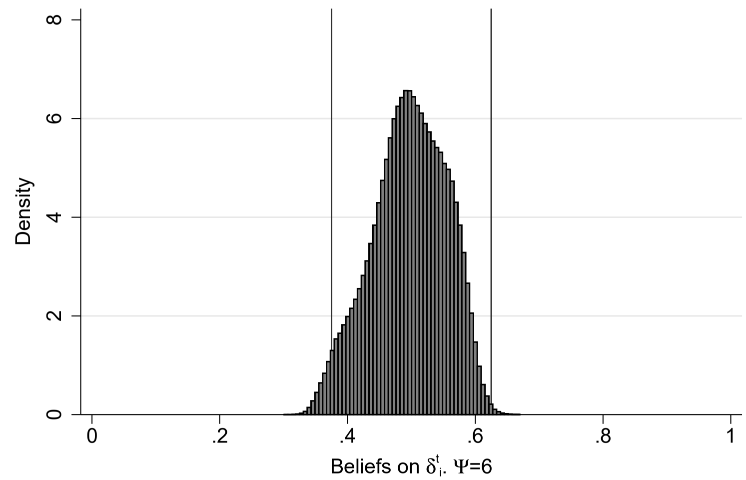
To determine convergence towards a stable state, I run 30 chains of simulations with randomly determined agents' priors. I choose 30 in order to provide a sample big enough to determine convergence, as described below. Then, every 50 generations, I discard the first half of observations for each chain, and find, for each chain, the standard deviation (s.d.) of the s.d. of beliefs and the CV. Then, I evaluate the ratio of the average within-chain s.d. and CV to the pooled s.d. and CV. Convergence is achieved if the squared root of both these ratios is less than 1.1, a rule of thumb common in the literature of MCMC. As I find that the simulation converge fairly quickly under the selected parameters, I further extend the number of simulated generations to more than three times what is needed for convergence. Then by taking three non-overlapping chains of each chain I multiply the set of chains used to estimate standard deviations of the patterns studied.

With the assumed $\rho = 0.5$, only 50 generations are needed for beliefs to converge.[20] Thus, I run 200 iterations and divide generations 100 to 199 in three groups to obtain a total of 90 chains (each with 33 generations). By having multiple chains, I can estimate standard deviations in all the simulation exercises. Figure V presents the resulting pooled distributions of beliefs in the stable states, as well as two lines delimiting the region of efficient choices described in the previous section. In the stable state, all the beliefs formed under very low levels of segregation are located within the region of efficient choices. Thus, as it was the case when all lineages start with the same beliefs, very low levels of segregation have no effect on effort choices. As segregation increases, distinct groups of agents emerge. When the segregation is low, the distribution is concentrated close to the actual values, though a second group of agents cluster around the low threshold of the region of efficient choices. Under high and very high levels of segregation, there are three clear groups of agents, with some under- (over-) estimating the returns to effort. When the level of segregation is high, still the majority lie in the region of efficient choices. But when segregation is very high, the great majority are underestimating the returns to effort.

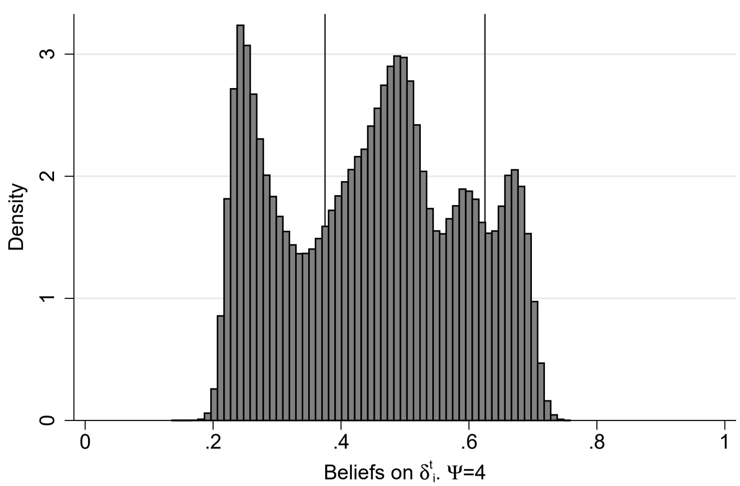
[20]Clearly the speed of convergence is related to the discounting parameter ρ. With a $\rho = .95$ convergence is achieved after 300 generations, on average.



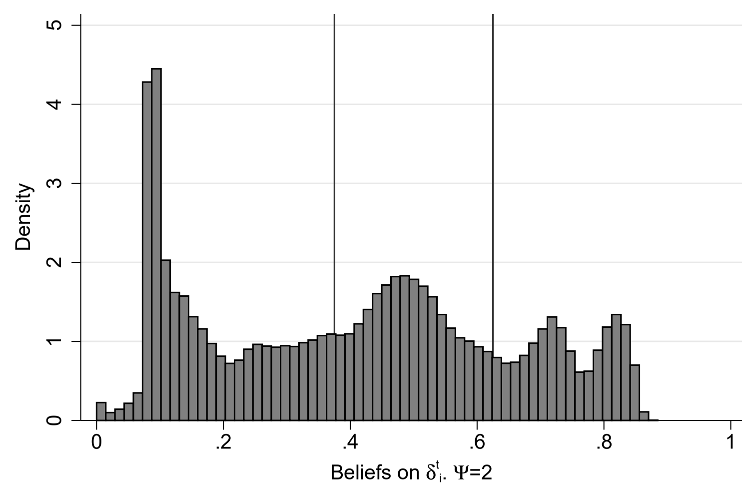
(A) $\psi = .8$



(B) $\psi = .6$



(C) $\psi = .4$



(D) $\psi = .2$

Figure V
Beliefs regarding δ in the stable state.

Beliefs and lineages experiences

Finding considerable differences in beliefs, I proceed to study what experiences lead different lineages to hold different beliefs. In this case, however, one must take into account that agents are only partially confident about the beliefs they receive from their parents. Therefore, I construct a measure of influence as follows. First, experiences of the generations in each chain are discounted by the same factor as agents are partially confident about their parent's beliefs $\rho = 0.5$. Then, discounted experiences are normalized to a number between zero and one by dividing each experience's discounted summation to the addition of all experiences discounted summations. That is, the influence of experience x is:

$$\text{Influence of experience}_x = \frac{\sum_{t=t_0}^{T} \rho^{(T-t)} \mathbb{1}^x[(e_i^t, \omega_i^t)]}{\sum_{y \in \mathbb{Y}} \sum_{t=t_0}^{T} \rho^{(T-t)} \mathbb{1}^y[(e_i^t, \omega_i^t)]}, \tag{10}$$

where $\mathbb{Y}$ denotes the set of all possible experiences and $\mathbb{1}^k[\cdot]$ is an indicator whether the experience lived was k. Notice that, under partial confidence, recent experiences have a larger effect on beliefs.

Table V presents the influence of different experiences on the formation of beliefs held by groups of lineages across the distribution of $\hat{\delta}$ under very low and very high segregation. Lineages are grouped according to the quintile their $\hat{\delta}$ is in the distribution of beliefs by the last generation of each chain. Average beliefs for each group are presented in the first column. The table also presents the average influence of different experiences for the whole population, and the average $\hat{\delta}$ each group of lineages holds by the last generation. Panel A presents the case of very low segregation. In this case, no lineage holds extreme beliefs so beliefs do not affect effort decisions. Therefore, two thirds of all the experiences are by agents who faced a low cost and thus exerted effort. This table tell us, for example, when segregation is low a typical lineage holding a very low belief regarding δ is relatively more influenced by experiences resulting from exerting no effort as 40% of their information was obtained when exerting no effort, compared to 33% of all agents. In particular, 34.4% of the information of the group holding a very low belief comes from experiences of exerting no effort and receiving a low income, and 5.6% from experiences of exerting no effort and receiving a high income. Additionally, 60.1% of their information comes from experiences exerting a high effort: 31.7% of times receiving a low income and 28.4% of times receiving a high income. Differences of beliefs, when beliefs do not affect effort decisions, are mainly driven by the number of times a lineage has drawn a low cost of effort and exerted effort. Because exerting effort increases the chances of receiving a high income, when segregation is very low the group holding the highest beliefs also have a higher average income.

TABLE V

Normalized measures of experiences' influence of groups with different levels of beliefs regarding δ.

Panel A – Very low segregation

	$\bar{\hat{\delta}}$	e_0, ω_l	e_0, ω_h	e_1, ω_l	e_1, ω_h
Very low belief	0.476	0.344	0.056	0.317	0.284
	(0.024)	(0.028)	(0.016)	(0.026)	(0.028)
Low belief	0.495	0.264	0.085	0.248	0.402
	(0.024)	(0.039)	(0.021)	(0.030)	(0.041)
Average belief	0.505	0.209	0.083	0.165	0.543
	(0.026)	(0.032)	(0.015)	(0.028)	(0.037)
High belief	0.510	0.232	0.070	0.148	0.550
	(0.026)	(0.031)	(0.017)	(0.021)	(0.036)
Very high belief	0.517	0.285	0.043	0.116	0.557
	(0.025)	(0.030)	(0.014)	(0.018)	(0.029)
Total	0.501	0.267	0.067	0.199	0.467
	(0.025)	(0.006)	(0.006)	(0.009)	(0.009)

Note: Standard deviations in parentheses.

Panel B – Very high segregation

	$\bar{\hat{\delta}}$	e_0, ω_l	e_0, ω_h	e_1, ω_l	e_1, ω_h
Very low belief	0.094	0.823	0.136	0.036	0.005
	(0.010)	(0.028)	(0.022)	(0.008)	(0.004)
Low belief	0.234	0.682	0.166	0.091	0.062
	(0.026)	(0.048)	(0.025)	(0.018)	(0.028)
Average belief	0.425	0.336	0.141	0.185	0.338
	(0.032)	(0.068)	(0.032)	(0.034)	(0.069)
High belief	0.555	0.200	0.078	0.196	0.527
	(0.032)	(0.023)	(0.024)	(0.030)	(0.037)
Very high belief	0.756	0.155	0.028	0.170	0.647
	(0.020)	(0.012)	(0.007)	(0.022)	(0.024)
Total	0.413	0.439	0.110	0.135	0.316
	(0.021)	(0.024)	(0.011)	(0.010)	(0.023)

Note: Standard deviations in parentheses.

TABLE VI

AVERAGE $\hat{\delta}^T$ BY MOBILITY EXPERIENCES WHEN $\psi = .6$.

t-1 \ t	ω_l	ω_h
ω_l	0.433	0.537
	(0.007)	(0.007)
ω_h	0.515	0.499
	(0.006)	(0.008)

Note: Standard deviations in parentheses.

When segregation is very high (Panel B) less than two thirds of all agents exert effort. This is the effect of a considerable number of agents holding extremely low $\hat{\delta}_i^t$s—typical of this environment. Agents holding low or very low $\hat{\delta}_i^t$s are almost never exerting effort when they have a low cost. Conversely, The group holding very high beliefs are exerting effort in most of recent generations. When segregation is very high there is a strong relationship between beliefs and the number of times an agent exerts effort, so beliefs become self-fulfilling in a sense. It is noteworthy that both exerting a high effort too often, as the group holding very high beliefs, or exerting no effort during the whole period, as the groups holding low and very low beliefs, is not optimal.

Beliefs and income mobility

Piketty (1995) points out to the central role of income mobility in shaping up individual beliefs. That is, political attitudes respond not only on an individual's income but also on their social background. For example, he mentions the fact that, while low-income individuals are generally more likely to vote for left-parties (which traditionally support higher levels of redistribution), low-income individuals with high-income parents are less likely to vote for left-wing parties than low-income individuals with low-income parents. As previously noted, individuals who care about the role of effort and luck in determining incomes will desire higher levels of redistribution if they believe effort matters less, and thus will be more likely to vote for a left-party. Here I study the effects of income mobility in beliefs when economies have reached a stable state. For simplicity, I focus on the case when segregation is very low ($\psi = .8$) so all agents are choosing optimally when to exert effort.

Table VI presents the average beliefs held by agents with high and low incomes by the level of their parent's income. Low-income agents with low-income parents hold the lowest estimates on the returns to effort ($\hat{\delta} = .43$), followed by high-income agents with high-income parents. Conversely, low-income agents with high-income parents hold the highest

TABLE VII

Detailed mobility esperiences and beliefs updating when $\psi = .6$.

Panel A – Average prior $\hat{\delta}^T$

t-1 \ t	ω_l		ω_h	
	$e=0$	$e=1$	$e=0$	$e=1$
ω_l				
$e=0$	0.480	0.500	0.481	0.500
	(0.014)	(0.006)	(0.015)	(0.006)
$e=1$	0.445	0.456	0.446	0.456
	(0.008)	(0.005)	(0.008)	(0.005)
ω_h				
$e=0$	0.473	0.476	0.474	0.476
	(0.009)	(0.008)	(0.010)	(0.007)
$e=1$	0.522	0.523	0.523	0.523
	(0.007)	(0.007)	(0.007)	(0.007)

Note: Standard deviations in parentheses.

Panel B – Average updated $\hat{\delta}^T$

t-1 \ t	ω_l		ω_h	
	$e=0$	$e=1$	$e=0$	$e=1$
ω_l				
$e=0$	0.443	0.427	0.499	0.570
	(0.011)	(0.006)	(0.008)	(0.006)
$e=1$	0.436	0.417	0.441	0.515
	(0.007)	(0.005)	(0.008)	(0.006)
ω_h				
$e=0$	0.495	0.426	0.463	0.491
	(0.008)	(0.007)	(0.009)	(0.007)
$e=1$	0.551	0.484	0.476	0.504
	(0.005)	(0.006)	(0.008)	(0.008)

Note: Standard deviations in parentheses.

Panel C – Change of $\hat{\delta}^T$ (in SD).

t-1 \ t	ω_l		ω_h	
	$e=0$	$e=1$	$e=0$	$e=1$
ω_l				
$e=0$	-0.693	-1.380	0.332	1.317
	(0.079)	(0.046)	(0.156)	(0.048)
$e=1$	-0.174	-0.731	-0.106	1.113
	(0.033)	(0.031)	(0.090)	(0.052)
ω_h				
$e=0$	0.415	-0.931	-0.206	0.289
	(0.101)	(0.061)	(0.082)	(0.045)
$e=1$	0.550	-0.730	-0.872	-0.362
	(0.076)	(0.053)	(0.058)	(0.038)

Note: Standard deviations in parentheses.

Panel D – % of mobility stories

t-1 \ t	ω_l		ω_h	
	$e=0$	$e=1$	$e=0$	$e=1$
ω_l				
$e=0$	0.082	0.052	0.020	0.122
	(0.007)	(0.002)	(0.002)	(0.003)
$e=1$	0.056	0.037	0.014	0.087
	(0.002)	(0.001)	(0.001)	(0.003)
ω_h				
$e=0$	0.019	0.014	0.005	0.032
	(0.001)	(0.001)	(0.001)	(0.002)
$e=1$	0.120	0.090	0.030	0.211
	(0.003)	(0.002)	(0.002)	(0.007)

Note: Standard deviations in parentheses.

estimates, preceded by low-income agents with high-income parents. On the assumption that support for redistribution is inversely related to the beliefs regarding δ, in general, our simulated economies follow a pattern somewhat consistent with the empirical regularities described by Piketty (1995). In particular, persistently low-income agents hold the lowest estimates on the returns to effort (thus, expected to be more supporting of redistribution and left wing parties), but parental income also matters. The only exception are the persistently high-income agents who do not have the highest estimates of the returns to effort. Yet, it is noteworthy that selfish motives are not included in the simulation as there is no redistribution.

The advantage of the simulations is that I can study more in detail the effects of mobility. In particular, there might be some differences between agents moving in the income distribution because they exert a different level of effort than their parent's or due to luck.

Table VII presents a detailed account of the process of beliefs' updating according to agents' and their parents' income and effort levels. Panel A describes the average priors. It says, for example, that parents who exerted effort and received a low income had an average prior of 0.445. Notice that parents receiving the unexpected income (low income when exerting effort or high income when exerting no effort) bequest lower beliefs than parents who received the expected incomes. On average, low-income parents hold lower estimates and high-income parents bequest considerably higher estimates, the highest belonging to high-income parents who exerted effort.

Panel B presents updated beliefs. Here it becomes clear how the forces behind a story of income mobility matter. Agents who move because of their choices (low-income agents who exerted no effort with high-income parents who exerted effort, and high-income agents who exerted effort with low-income parents who exerted no effort) are the ones holding the highest estimates on the returns to effort. This is because, as shown in Panel C, downwardly moving agents by choices are the only ones increasing beliefs (by half a standard deviation) of high income parents who exerted effort (who hold the highest priors to begin with). Upwardly mobile agents, on the contrary, hold the highest beliefs because they increase their priors by almost one and a half standard deviations. Note that, as shown in Table VI, the upwardly mobile are the ones holding the highest beliefs on average. But this is because the most frequent way to move upwardly is by exerting effort when the parent did not, as shown in Panel D. Other economies might have different general patters depending on the relative frequencies of experiences, which is related to the parameters θ and δ.

V. Concluding remarks

This paper studies the extent to which social segregation can explain significant and persistent differences in beliefs regarding the role of effort in determining incomes. First I show some evidence in this regard, at least for the U.S. And then I present a model explaining why individuals, who can learn from the experiences of others, can persistently disagree about the role of effort in determining incomes. In the model, agents of different lineages learn naively in a seemingly optimal way from the observation of the experiences of others in their networks in addition to their own experience. Networks are biased towards their own experience due to social segregation.

Simulating economies of agents learning in this way, I show that social naivete when learning and social segregation can explain significant disagreements regarding the role of effort in increasing one's probability of receiving a high-income. Moreover, the model predicts that under high levels of segregation some agents may take sub-optimal choices while drifting towards incorrect beliefs, and that the level of disagreement between agents increases with the level of social segregation.

It is noteworthy that in the real world holding an extreme belief may have implications beyond inefficient effort decision making. Extreme beliefs can also affect how individuals interpret the actions of others. For instance, an individual holding the belief that δ is too high, so she exerts effort even when facing a high cost, could question the motives of another individual exerting no effort. She may think, for example, that either (i) the other individual has not been able to solve the problem, or (ii) the cost of effort is sufficiently high—though still unobserved by her lineage—so it was optimal for the other individual to choose differently. If an individual considers any of these two cases as the reason behind the other agent's choice of exerting no effort, they might consider that the other individual is less deserving of their income. The converse is true for agents holding extremely low estimates on the returns to effort. Therefore, holding extreme beliefs can also affect an individual's willingness to redistribute income.

On the assumption that widespread political disagreement is not beneficial to the economy, my findings suggest we should be concerned about increasing segregation as this leads to accruing disagreements over beliefs which are fundamental to policy questions. Specifically, under an increasing income segregation one can expect higher disagreements with respect to the relative roles of effort and luck in determining incomes which is expected to affect individual choices as well as debates about the optimal degree of redistribution in a society and the efficiency of governments.

References

Acemoglu, D., A. Nedic, and A. Ozdaglar (2008). Convergence of rule-of-thumb learning rules in social networks. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pp. 1714–1720.

Akerlof, G. A. (1997). Social distance and social decisions. *Econometrica 65*(5), 1005–1027.

Alesina, A. F. and G.-M. Angeletos (2005). Fairness and Redistribution. *American Economic Review 95*(4), 960–980.

Alesina, A. F. and P. Giuliano (2010). Preferences for redistribution. In J. Benhabib, A. Bisin, and M. O. Jackson (Eds.), *Handbook of Social Economics*, Volume 1A, Chapter 4, pp. 93–131. Cambridge, MA, USA: Elsevier B.V.

Alesina, A. F., E. L. Glaeser, and B. Sacerdote (2001). Why Doesn't the United States Have a European-Style Welfare State? *Brookings Papers on Economic Activity* (2), 187–277.

Alesina, A. F. and G. Tabellini (1990). A Positive Theory of Fiscal Deficits and Government Debt. *The Review of Economic Studies 57*(3), 403–414.

American Communities Project (2011). Income Segregation. Dataset.

Arifovic, J., B. C. Eaton, and G. Walker (2015). The coevolution of beliefs and networks. *Journal of Economic Behavior & Organization 120*, 46–63.

Azzimonti, M. (2011). Barriers to investment in polarized societies. *American Economic Review 101*(5), 2182–2204.

Bala, V. and S. Goyal (1998). Learning from neighbors. *Review of Economic Studies 65*(3), 1–41.

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

Barber, M. and N. Mccarty (2013). Causes and Consequences of Polarization. In J. Mansbridge and C. J. Martin (Eds.), *Task Force on Negotiating Agreement in Politics*, Chapter 2, pp. 19–53. Washington, DC, USA: American Political Science Association.

Benabou, R. and J. Tirole (2006). Belief in a just world and redistributive politics. *Quarterly Journal of Economics 121*(2), 699–746.

Binder, S. A. (1999). The Dynamics of Legislative Gridlock, 1947-96. *The American Political Science Review 93*(3), 519–533.

Bischoff, K. and S. F. Reardon (2014). Residential Segregation by Income, 1970-2009. In J. R. Logan (Ed.), *Diversity and Disparities America enters a new century*, Chapter 7, pp. 208–234. New York, NY, USA: Russell Sage Foundation.

Bisin, A. and T. Verdier (2005). Cultural Transmission.

Blair, J. and M. G. Lacy (2000). Statistics of Ordinal Variation. *Sociological Methods & Research 28*(3), 251–280.

Clark, T. S. and D. a. Linzer (2015). Should I Use Fixed or Random Effects? *Political Science Research and Methods 3*(2), 399–408.

Cutler, D. M. and E. L. Glaeser (1997). Are ghettos good or bad? *Quarterly Journal of Economics 112*(3), 827–872.

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

Di Tella, R. and J. Dubra (2008). Crime and punishment in the ”American Dream”. *Journal of Public Economics 92*(7), 1564–1584.

Dillenberger, D. and P. Sadowski (2012). Ashamed to be selfish. *Theoretical Economics 7*(1), 99–124.

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

Fong, C. M. (2001). Social preferences, self-interest, and the demand for redistribution. *Journal of Public Economics 82*(2), 225–246.

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

Galor, O. and J. Zeira (1993). Income distribution and macroeconomics. *The Review of Economic Studies 60*(1), 35–52.

Gelman, A. and K. Shirley (2011). Inference from simulations and monitoring convergence. *Handbook of Markov Chain Monte Carlo*, 163–174.

Harris, P. L. and K. H. Corriveau (2011). Young children's selective trust in informants. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences 366*(1567), 1179–87.

Hirschman, A. O. (1964). The Paternity of an Index. *The American Economic Review 54*(5), 761.

Iceland, J. and R. Wilkes (2006). Does socioeconomic status matter? Race, class, and residential segregation. *Social Problems 53*(2), 248–273.

Jackson, M. O. and A. Wolinsky (1996). A Strategic Model of Social and Economic Networks. *Journal of Economic Theory 71*(1), 44–74.

Jensen, R. T. (2010). The (perceived) returns to education and the demand for schooling. *Quarterly Journal of Economics 125*(2), 515–548.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review 93*(5), 1449–1475.

Kahneman, D. (2013). *Thinking, Fast and Slow* (Reprint ed.). Farrar, Straus and Giroux.

Massey, D. S., J. Rothwell, and T. Domina (2009). The Changing Bases of Segregation in the United States. *The Annals of the American Academy of Political and Social Science 626*(1), 74–90.

Mcfadden, D. (2011, aug). 2011 Lindau Nobel Laureate Meetings: Panel Discussion.

Meltzer, A. H. and S. F. Richard (1981). A rational theory of the size of government. *The Journal of Political Economy 89*(5), 914–927.

Mills, C. M. (2013). Knowing when to doubt: developing a critical stance when learning from others. *Developmental psychology 49*(3), 404–18.

Mobius, M. and T. Rosenblat (2014). Social Learning in Economics. *Annual Review of Economics 6*(1), 827–847.

National Center for Education Statistics (2016). High School Longitudinal Study of 2009-2013 [United States]. Dataset. Retreived from https://nces.ed.gov/surveys/hsls09/.

Page, L. and D. G. Goldstein (2016). The role of subjective beliefs in preferences for redistribution. *Social Choice and Welfare*, 1–37.

Pew Research Center for the People & the Press (2009). Trends in Political Values and Core Attitudes, 1987-2009. Dataset. Retreived from http://www.thearda.com/Archive/Files/Descriptions/VAL8709.asp.

Piketty, T. (1995). Social mobility and redistributive politics. *Quarterly Journal of Economics 110*(3), 551–584.

Rabin, M. (1995). Moral preferences, moral constraints, and self-serving biases. Working Paper.

Reardon, S. F. (2009). Measures of ordinal segregation. In Y. Flückiger, S. F. Reardon, and J. Silber (Eds.), *Occupational and Residential Segregation (Research on Economic Inequality)*, Volume 17, pp. 129–155. Emerald Group Publishing Limited.

Reardon, S. F. and K. Bischoff (2011). Income inequality and income segregation. *American Journal of Sociology 116*(4), 1092–153.

Reardon, S. F., G. Firebaugh, D. O'Sullivan, and S. Matthews (2006). A New Approach to Measuring Socio-Spatial Economic Segregation. In *Paper Prepared for the 29th General Conference of The International Association for Research in Income and Wealth*, Volume 9931, pp. 1–48.

Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2015). Integrated Public Use Microdata Series: Version 6.0. Machine-readable database.

Scheinkman, J. A. (2008). Social interactions (theory). In S. N. Durlauf and L. E. Blume (Eds.), *The New Palgrave Dictionary of Economics* (2nd ed.). Palgrave Macmillan.

Schelling, T. C. (1971). Dynamic models of segregation. *The Journal of Mathematical Sociology 1*(2), 143–186.

Tversky, A. and D. Kahneman (1971). Belief in The Law of Small Numbers. *Psychological Bulletin 76*(2), 105–110.

Tversky, A. and D. Kahneman (1974). Judgment under Uncertainty: Heuristics and Biases. *Science 185*(4157), 1124–31.

Vickrey, W. (1945). Measuring marginal utility by reactions to risk. *Econometrica 13*(4), 319–333.

World Values Survey Association (2015). World Values Survey 1981-2014 Longitudinal Aggregate v.20150418. Dataset. Retrieved from http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp.

Appendix A. Income-based segregation and differences in beliefs about the role of effort in the U.S.

In this appendix, I study the relationship between income-based segregation and the differences in individual beliefs regarding the role of effort in determining success, in the U.S. at the metropolitan area level. I focus on metropolitan areas as this the most disaggregated geographical level available in my data where individuals are likely to meet others.[21]

I use a measure of Income Segregation provided by the American Communities Project (2011). This dataset includes several measures of segregation by family income for every decade between the 1970s and the 2010s. In particular, it includes measures of overall income segregation, poverty segregation and affluence segregation for 380 metropolitan areas. The income segregation index used by them is the rank-order information theory index by Reardon et al. (2006). In short, this is a measure proportional to the amount of population that should be relocated in order to have an income distribution within neighborhoods equal to the income distribution of the metropolitan area they are in. Poverty and affluence segregation refer to the distribution between the lower 10% and the rest, and the lower 90% and the rest respectively.

First I study the correlation between these variables and find a significant relationship between income segregation and the level of differences in beliefs. Given that income inequality might be explaining this relationship, I also study to what extent this relationship may be explained by the correlation between income inequality and income segregation. I find that there is evidence suggestive of a role for both income inequality and income segregation in shaping beliefs.

A.1 Data

Dispersion of beliefs

In the Trends in Political Values and Core Attitudes 1987-2009 surveys by the Pew Research Center for the People & the Press (2009), survey respondents were asked to rate their level of agreement with the statement that "[h]ard work offers little guarantee of success" (FEELWORK).[22] To express their level of agreement, survey respondents used a likert scale with four options ranging from "completely agree" to "completely disagree." Based on the samples of survey responses at the metropolitan area level provided by this dataset, I estimate the level of differences in beliefs.

[21] As many theories of social contact are constructed on the idea of neighborhoods, optimally I would like to study this relationship at the neighborhood level. However, to the best of my knowledge sufficiently big samples of surveys at this level do not exist. Metropolitan Areas are nevertheless, still suitable for my analysis. According to the U.S. Census, generally speaking "a metropolitan area (MA) is a core area containing a large population nucleus, together with adjacent communities that have a high degree of economic and social integration with that core." Thus, there is always some likelihood of individuals meeting others from their MAs.

[22] Additionally, a portion of respondents were also asked to rate "[m]any people today think that they can get ahead without working hard and making sacrifices." However, sample sizes within metropolitan areas of this question are too small to provide any reliable measure of dispersion. Some years also included statements related to the role of the government in helping people in need and beliefs of why people may remain being poor, with the same problem of sample sizes.

When estimating the difference in beliefs of each metro area, there are two important considerations. First, beliefs were measured using a Likert scale. Unlike continuous variables, ordinal variables—like the ones provided by Likert scales—do not have a clear measure of dispersion. Namely, Likert scales reflect the position of an unobservable underlying attitude in an unknown distribution relative to some unknown thresholds and only provide an order, but no information about the expected "difference" in the underlying attitudes, between options (e.g. between "completely agree" and "disagree"). I consider one parametric estimation of the differences between categories to estimate a standard deviation, and the non parametric measure developed by sociologists of Ordinal Variation (OV) (Reardon, 2009; Blair and Lacy, 2000).

Second, the survey is designed to be representative at the national level, not the metropolitan area level. Therefore, while the samples at each metro area are expected to be random, the number of people surveyed in each metro area varies significantly and can be in some cases very small. This might be problematic as, for example, by chance it is more likely to observe highly dispersed samples when there are few survey respondents, even if the overall population is not so dispersed. Also, this might be a concern, if the actual measure of dispersion of beliefs is somehow related to the number of surveys conducted in each metro area. I propose a method of bootstrapping to account for any potential effects of differences in sample sizes on the measures of dispersion.

Measuring the amount of differences in beliefs when beliefs are expressed using Likert scales. I consider the index of Ordinal Variation (OV) developed by sociologists.[23] The index of OV is a non parametric method developed by sociologists to avoid making any assumptions about the differences between categories. It relies in the empirical cumulative distribution function and the number of possible answers. OV is measured as:

$$OV = \frac{1}{K-1} \sum_{j=1}^{K-1} 4F[j](1 - F[j]) \tag{A.1}$$

[23]Other alternatives are (i) the standard deviation of categories themselves, (ii) the complement the Hirschman-Herfindahl Index (i.e. treating the variable as nominal), and (iii) dichotomizing the answers. Calculating the standard deviation of categories themselves relies on the unlikely assumption that the expected difference in the underlying factor between categories is constant—or almost constant. The complement of the Hirschman-Herfindal Index (Hirschman, 1964, CHHI hereafter) treats the variable as nominal and measures the concentration of answers in categories, providing one measure of "consensus." If all survey respondents choose the same option then the index takes a value of 1, and its complement a value of 0. Yet, the CHHI might or might not be an improvement over parametric estimates of variance. On the one hand, the CHHI makes no assumption on the difference between categories, but on the other hand it completely disregards the ordinal nature of the data. Therefore, the dispersion in a group where all but one respondents are in "complete disagreement" and one only "disagrees" would be the same as the dispersion in a group where all but one respondents are in "complete disagreement" and one "completely agrees." Finally, If categories can be grouped into two qualitatively different groups, e.g. agreement and disagreement, one can dichotomize the measure of beliefs under a dummy variable. The advantage is that, in the binary case there is a well defined measure of standard deviation. The problem is that valuable information (variation) is lost in the process and the variance of the resulting variable might be very small if one group is far more common than the other. I do not consider the first alternative given its overly restrictive assumptions, though the results are similar to using the "estimated positions." Also, I do not consider alternatives two and three as they show no correlation with income inequality contrary to what most of the literature suggests.

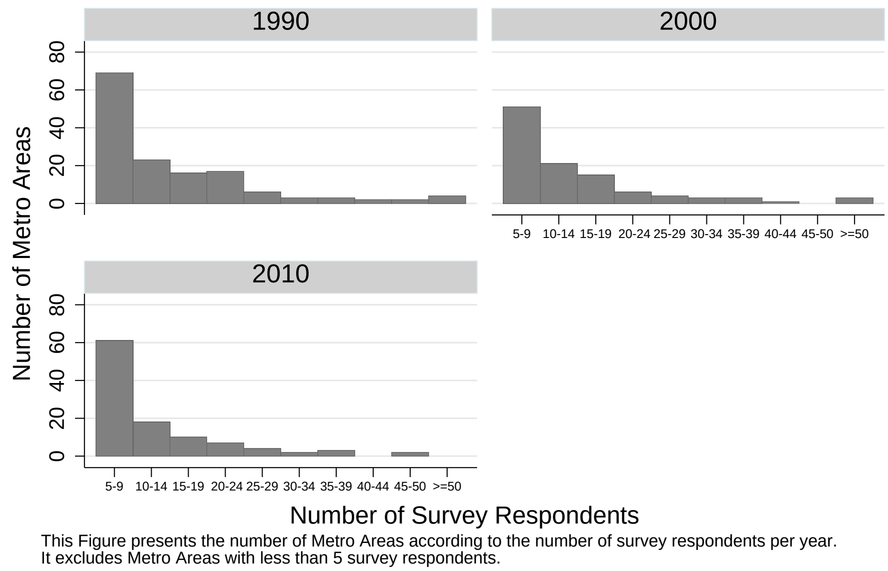


Figure A.1
Distribution of metropolitan areas according to the number of survey respondents restricted to areas with 5 or more respondents.

where K is the number of options and $F[j]$ is the empirical CDF. It takes a maximum value of 1 when half of respondents choose the first option and half choose the last option. It takes a minimum value of 0 when all respondents are in one category. This index is a measure of concentration, similar to the Hirschman-Herfindahl Index. Yet, unlike the HHI, according to this index an even split between disagree and agree is different than an even split between strongly agree and strongly disagree. For example, in our case with four options, an even split between categories 1 and 2 corresponds to an OV of $4 * 0.5 * (1 - 0.5)/3 + 4 * 1 * (1 - 1)/3 + 0 = 0.33$, but the OV of an even split between 1 and 3$=$ $4 * 0.5 * (1 - 0.5)/3 + 4 * 0.5 * (1 - 0.5)/3 + 4 * 1 * (1 - 1)/3 = 0.66$. While measures based on the CDF, such as the OV, are frequently preferred, they also exhibit the feature that even splits between "completely disagree" and "disagree" will be considered equally dispersed to even splits between "disagree" and "agree."

Accounting for differences in sample sizes between metro areas. Equally important to choosing the measure of dispersion is dealing with the differences in sample sizes between metropolitan areas. Differences in sample sizes might matter as they are related to the precision of the estimates of the actual amount of differences in beliefs within the metro area, and small samples might even bias some of the estimates of dispersion. In particular, small samples bias upwardly the estimates of the standard deviation, but not the measure of Ordinal Variation.[24] Yet, when there are considerable differences in the samples sizes of

[24] A simple exercise to observe the bias is to estimate the standard deviation of samples equally divided between two different numbers. You can notice how the estimated standard deviation increases as the sample

each metro area, more precise estimates of the differences within metro areas bear the cost of decreasing the number of metro areas in the analysis. Figure A.1 presents the distribution of metropolitan areas according to the number of survey respondents for metro areas with at least 5 survey respondents.[25] A majority of metropolitan areas have between 5 and 10 survey responses. What is more, some thresholds seemed to have been targeted by the survey enumerators as the number of metro areas decreases significantly after 10 and 20 survey responses.

For regression analyses I use a method of bootstrapping (explained below) that equates the number of survey responses when estimating the amount of difference in beliefs within metropolitan areas. Taking into account the number of survey responses in the different metro areas, I consider 5, 10 or 20 survey responses when running the bootstraps.

Income Inequality

An additional exercise controls for income inequality. Measures of per-capita income inequality are estimated using the samples available at the USA IPUMS by Ruggles et al. (2015). In particular, I measure the ratio of the 90th to the 10th percentile at the metropolitan area level of the IPUMS 1990 5% State sample, the IPUMS 2000 5% State sample, and the IPUMS 2010 1% ACS sample. Per-capita income is defined as the ratio between the household income (HHINCOME) and the number of persons in the household (NUMPREC).[26]

A.2 Empirical Strategy

Simple correlation

To begin, I study whether there is any correlation between income segregation and the amount of differences in beliefs, as measured by the OV, using a weighted analysis. Figure I presents the relationship between the income segregation index and the OV of survey answers in metropolitan areas with at least 5 survey respondents. To take into account for differences in sample sizes (precision), the scatter plot and the linear regression are weighted by the number of survey respondents. The figure suggest a positive and significant relationship between the level of income segregation and the dispersion of survey responses regarding the importance of effort in guaranteeing success. As a matter of fact, in any given year and with more restrictive samples (i.e. with at least 10 or 20 survey responses) weighted regressions—like the one depicted in the figure—indicate a positive and significant relationship between income segregation and the amount of differences in beliefs as measured by the OV.[27]

I also analyze bootstrapped regressions that keep the sample sizes equal when estimating the amount of differences in beliefs. Each bootstrap is made as follows: (i) select random

size drops below 100.

[25] Little is expected to be learned from any measure of dispersion based on less than 5 survey responses, so I only consider metro areas with 5 or more respondents.

[26] As recommended by IPUMS, when calculating income inequality I use Household weights (HHWT).

[27] Considering only the metro areas with at least 5 survey responses, there is information of income segregation and survey responses for 145 metropolitan areas in 1990, 107 in 2000, and 107 in 2010. Regressions not presented here but are available on demand.

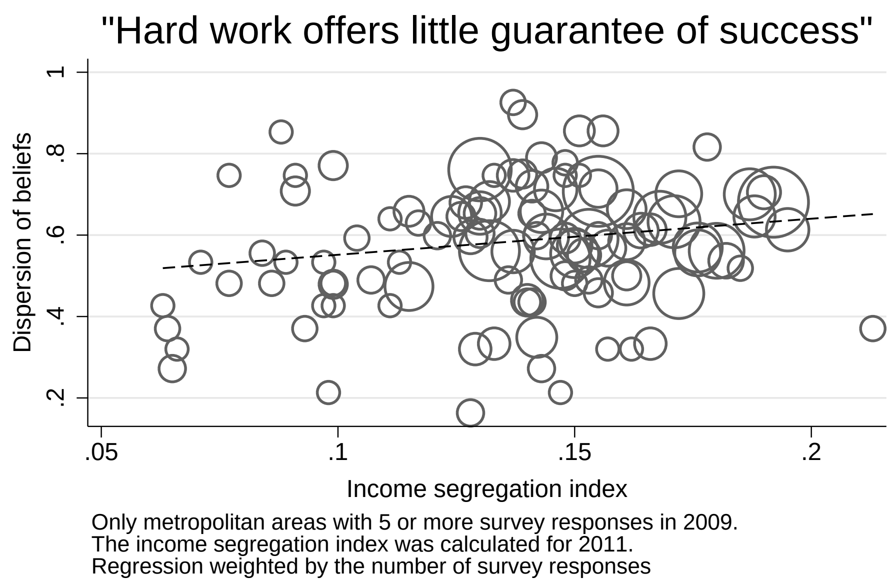


Figure A.2
Relationship between income segregation and the OV of beliefs regarding the importance of effort in guaranteeing success (5 or more respondents).

samples of a given n number of respondents without repetition for each metropolitan area, (ii) measure the level of dispersion within each sample, and (iii) regress, at the metropolitan area level, the resulting dispersion on the variables of interest. Every regression is bootstrapped a 1,000 times. The results are presented in Table A.1. Reported bootstraped p-values correspond to the percentile in the distribution of estimates where zero, or the first estimate with sign opposite to the mean, is found. There is also the possibility that no switching point is found, i.e. all estimates are positive or negative, in which case I define as having a p-value$<$.01. Bootstrapped estimates of pooled regressions indicate a positive and significant correlation between income segregation and the measure of OV, for all sample sizes. The relationship is, however, only significant for sample sizes of 20 when the amount of differences in beliefs is measured as the standard deviation of estimated positions.

Accounting for metropolitan area effects

A question, however, is whether differences in the dispersion of beliefs can be explained by unobserved characteristics of the metropolitan areas. Having random samples in multiple years (1990, 2000 and 2010) for multiple metro areas allows considering this possibility by using models for panel data. These are in the form of:

$$D_{i,t} = \alpha + \beta S_{i,t} + u_{i,t} + \epsilon_{i,t} \tag{A.2}$$

TABLE A.1

BOOTSTRAPPED RANDOM EFFECTS ESTIMATIONS (1,000 TIMES).

	Correlations				Both variables			
	Income Segregation		Ratio p90:p10		Income Segregation		Ratio p90:p10	
Dispersion (Ordinal Variation)	Mean β	B. p-val	Mean γ	B. p-val	Mean β	B. p-val	Mean γ	B. p-val
Ordinal Variation								
n=5	0.383	0.06	-0.001	0.38	0.205	0.24	-0.002	0.30
n=10	0.498	0.02	0.008	0.04	0.172	0.30	0.003	0.08
n=20	0.677	0.02	0.002	0.13	0.878	0.03	0.000	0.45

Notes: This table presents mean estimates of 1,000 bootstrapped regressions. In each bootstrap, samples sizes are 5, 10 or 20 survey respondents. Bootstrapped p-values are the percentile in the distribution of estimates where zero, or the first estimate with different sign of the mean, is found. In the case of finding no switching point, a p-value of 0 is assigned.

where $D_{i,t}$ is the measure of dispersion, $S_{i,t}$ is the level of segregation, $u_{i,t}$ is an unobserved characteristic of the metropolitan area, and $\epsilon_{i,t}$ is an strictly exogenous error term. The $u_{i,t}$s can be assumed to be constant within metro areas over time as in the fixed effects estimations (i.e. $u_{i,t} = u_i \forall i$), in which case they can be directly estimated (one for each metropolitan area). Fixed effects estimations allow for any correlation between $u_{i,t}$ and the idiosyncratic error $\epsilon_{i,t}$. Alternatively, one can assume that there is no correlation between $u_{i,t}$ and the idiosyncratic error $\epsilon_{i,t}$ ans use a random effects estimation. By assuming the $u_{i,t}$ follow a unique distribution, estimates of random effects are more efficient. However, they can be biased if there is some correlation between the level of segregation and the metropolitan areas effects.

Clearly, in most cases one would estimate a model of fixed effects as it allows for any type of correlation between the metro areas effects and our variables of interest. However, under certain conditions the more efficient model of random effects can be preferred. In particular, random effects are preferable when the number of observations per unit is low and either the dependent or the independent variables are "slow moving or sluggish" (Clark and Linzer, 2015). Changes in residential income segregation are very slow. Even if our observations are separated for a decade at a time, the within standard deviation is less than 25% of the overall standard deviation, which is small relative to the mean level of segregation to begin with. Therefore, for the analysis I estimate models of random effects.[28]

A.3 RESULTS

The first two columns of Table A.1 summarizes the results of bootstrapped random effects estimations of the effects of income segregation in the difference of beliefs. Results using the standard deviation of EP and the OV are reported for sample sizes $n = 5$, 10 or 20. Generally, there is a positive and significant relationship between the level of income segregation and the different measures of dispersion. This relationship is not significant only when the sample size is 5, which can reflect the measurement error introduced by very small samples.

[28]Estimates of fixed effects show no significant effects of income segregation nor income inequality.

Income inequality as an alternative explanation

It has to be acknowledged that the period of time under study is one of increasing income inequality. Dispersion of beliefs within societies may be due to selfish motives. Selfishness dictates that the rich desire lower levels of redistribution for which they benefit less than the poor (Meltzer and Richard, 1981). Indeed, the typical rich prefer less redistribution (Fong, 2001). However, individuals dislike feeling they are being selfish (Dillenberger and Sadowski, 2012). Instead, rich individuals may rationalize their preference for low redistribution by adjusting their beliefs regarding the role of effort and luck similar to the process described by Benabou and Tirole (2006). In that case, if income inequality increases—as it did during the period studied—selfish motives might explain an increasing dispersion of beliefs. Therefore, I also study the effect of controlling for income inequality when analyzing the relationship between income segregation and the dispersion of beliefs. Income inequality is measured as the Gini coefficient of a collection of samples by Ruggles et al. (2015). In the estimations, the coefficient for the added variable Gini ($G_{i,t}$) is γ.[29]

Columns 3 and 4 of Table A.1 presents mean bootstrapped coefficients for regressions including only the Gini coefficient, while columns 5 and 6 presents regressions including both variables. Income inequality by itself is always positively correlated with the dispersion of beliefs. When both income segregation and income inequality are included it is more common to find only one of the variables having a positive and significant effect. This might result from the high correlation that exists between the two measures. For $n = 5$ (the most imprecise) or $n = 20$ (the smaller sample size), the variable that is significant is the level of income inequality. Conversely, when $n = 10$ (the case with the best balance between precision and sample size) segregation has a positive and significant effect. It is noteworthy that with a sample size of 10, both income segregation and income inequality have a positive and significant effect when the amount of differences in beliefs is measured as the s.d. of the EP.

[29] There is information of income segregation, income inequality and at least 5 survey responses for 101 metropolitan areas in 1990, 80 in 2000, and 86 in 2010.

Appendix B. Analytical solution when beliefs do not affect effort choices

This appendix studies how beliefs evolve when effort is independent of beliefs and beliefs are perfectly transmitted. In the simulations this is the case, for example, when segregation is very low.

As it was previously assumed, agent's beliefs are described by beta distributions and the likelihood observed in their own experience and in others is described by a binomial distribution. In this case agents update beliefs regarding the probability of receiving a high income, for each level of effort $e \in \{0, 1\}$, according to:

$$\boldsymbol{\beta}^t_{i,e}\left(a^t_{i,e}, b^t_{i,e}\right) \propto B^{t-1}_{i,e}(n^t_{i,e}, p^t_{i,e}) \times \boldsymbol{\beta}^{t-1}_{i,e}(a^{t-1}_{i,e}, b^{t-1}_{i,e}) \tag{B.1}$$
$$\text{where}$$
$$a^t_{i,e} = a^{t-1}_{i,e} + p^{t-1}_{i,e} n^{t-1}_{i,e} \text{ and } b^t_{i,e} = b^{t-1}_{i,e} + (1 - p^{t-1}_{i,e}) n^{t-1}_{i,e}$$

The objective of this section is the study of long-run beliefs when effort choices are not affected by beliefs. Given that beliefs are perfectly transmitted, beliefs do converge to a unique value. As our object of interest the value to which they converge, and on the long-run all lineages will live similar experiences histories due to the Law of Large Numbers, the individual index is dropped. For example, let's start by finding what are the number of high and low incomes (parameters a^T_1 and b^T_1) an agent should observe on agents exerting effort (either themselves or others) after a sufficiently high T number of periods. That is,

$$\begin{aligned}
a^T_1 = \; & \underbrace{\lambda(\theta+\delta)T}_{\text{Times } e=1 \wedge \omega=h;} \; \underbrace{\lambda(\theta+\delta)N}_{\text{Agents with } e=1 \wedge \omega=h;} \; \underbrace{1}_{\text{Proportion met}} \\
+ \; & \underbrace{(1-\lambda)\theta T}_{\text{Times } e=0 \wedge \omega=h;} \; \underbrace{\lambda(\theta+\delta)N}_{\text{Agents with } e=1 \wedge \omega=h;} \; \underbrace{\psi}_{\text{Proportion met}\cdots} \\
& + \lambda(1-\theta-\delta)T\lambda(\theta+\delta)N\psi \\
& + (1-\lambda)(1-\theta)T\lambda(\theta+\delta)N\psi^2 \\
& = T\lambda(\theta+\delta)N\left(\lambda(\theta+\delta) + (1-\lambda)\theta\psi + \lambda(1-\theta-\delta)\psi + (1-\lambda)(1-\theta)\psi^2\right)
\end{aligned} \tag{B.2}$$

One can rewrite Equation B.2 as:

$$a^T_1 = T\lambda(\theta+\delta)N\Big[1 - (1-\psi)\Big[\lambda(1-\theta-\delta) + (1-\lambda)\Big(1 + \psi(1-\theta)\Big)\Big]\Big] \tag{B.3}$$

which shows the loss of sample resulting from segregation and the frequency of different experiences lived.

Similarly,

$$b^T_1 = T\lambda(1-\theta-\delta)N\Big[1 - (1-\psi)\Big[\lambda(\theta+\delta) + (1-\lambda)\Big(1 + \psi\theta\Big)\Big]\Big] \tag{B.4}$$

Therefore,

$$
\begin{aligned}
a_1^T + b_1^T &= T\lambda(\theta+\delta)N\Big[1-(1-\psi)\Big[\lambda(1-\theta-\delta)+(1-\lambda)\Big(1+\psi(1-\theta)\Big)\Big]\Big] \\
&\quad + T\lambda(1-\theta-\delta)N\Big[1-(1-\psi)\Big[\lambda(\theta+\delta)+(1-\lambda)\Big(1+\psi\theta\Big)\Big]\Big] \\
&= T\lambda N\Big[1-(1-\psi)\Big[\lambda 2(\theta+\delta)(1-\theta-\delta) \\
&\quad + (1-\lambda)\Big(1+\psi((\theta+\delta)(1-\theta)+\theta(1-\theta-\delta))\Big)\Big]\Big]
\end{aligned}
\tag{B.5}
$$

Hence, the mean belief regarding the probability of a high income when exerting effort, after T periods, will be:

$$
\frac{a_1^T}{a_1^T+b_1^T} = \frac{(\theta+\delta)\Big[1-(1-\psi)\Big[\lambda(1-\theta-\delta)+(1-\lambda)\Big(1+\psi(1-\theta)\Big)\Big]\Big]}{\Big[1-(1-\psi)\Big[\lambda 2(\theta+\delta)(1-\theta-\delta)+(1-\lambda)\Big(1+\psi((\theta+\delta)(1-\theta)+\theta(1-\theta-\delta))\Big)\Big]\Big]}
\tag{B.6}
$$

similarly,

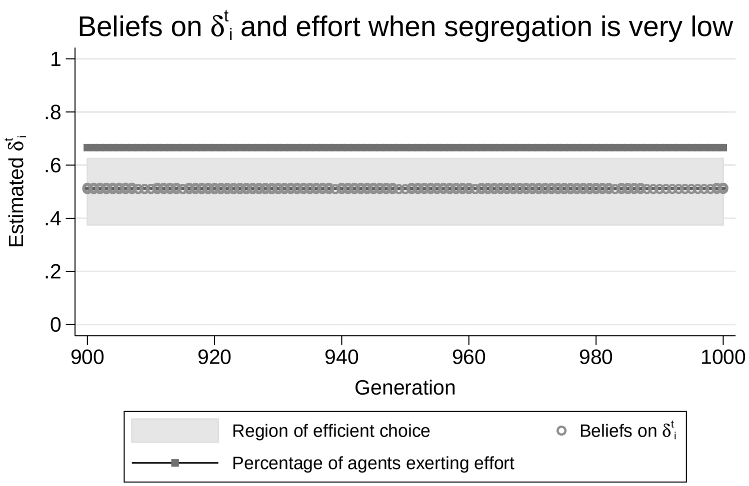
$$
\frac{a_0^T}{a_0^T+b_0^T} = \frac{\theta\Big[1-(1-\psi)\Big[\lambda\Big(1+\psi(1-\theta-\delta)\Big)+(1-\lambda)(1-\theta)\Big]\Big]}{\Big[1-(1-\psi)\Big[\lambda\Big(1+\psi((\theta+\delta)(1-\theta)+\theta(1-\theta-\delta))\Big)+(1-\lambda)2\theta(1-\theta)\Big]\Big]}
\tag{B.7}
$$

These expressions allow for studying the bias in $\hat{\theta}^T$ and $\hat{\theta}^T+\hat{\delta}^T$. In particular, $(\theta+\delta) > 1/2$ will introduce an upward bias in $\hat{\theta}^T+\hat{\delta}^T$ as will $\theta > 1/2$. If the chances of obtaining a high income are less than one half in one case, and greater than one half in the other, as we assumed in our simulation, the total bias will depend on which effect dominates. Note, however, that the probability of receiving a high income when exerting effort is expected to matter the most in this case as the term depending on θ is multiplied by ψ. Similarly, $\hat{\theta}^T$ will be upwardly biased if $(\theta+\delta) > 1/2$ or $\theta > 1/2$.

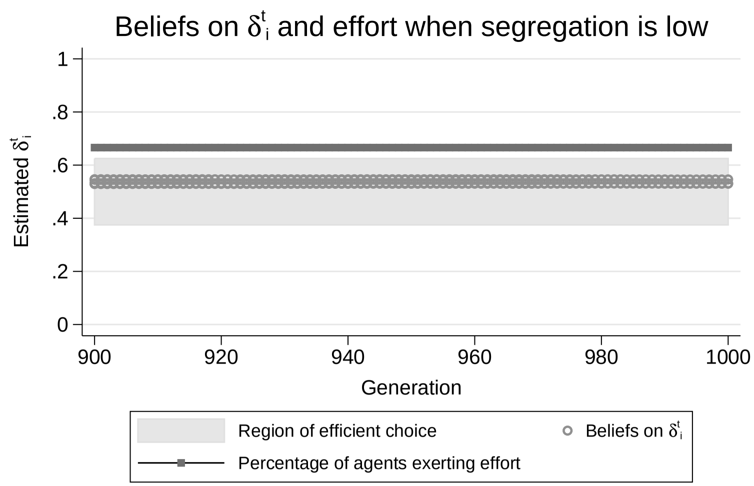
Appendix C. Long-run Beliefs under Perfect Intergenerational Transmission of Beliefs

Figure C.1 presents long-run beliefs with respect to the returns to effort when the transmission of beliefs is perfect, that is when the young are not skeptical about the beliefs they receive. If segregation is very low or low, agents' beliefs converge to a unique estimation, that nevertheless depends on the overall level of segregation. Namely, the beliefs of agents in the low segregated societies are somewhat higher than the estimates of very low segregated societies.

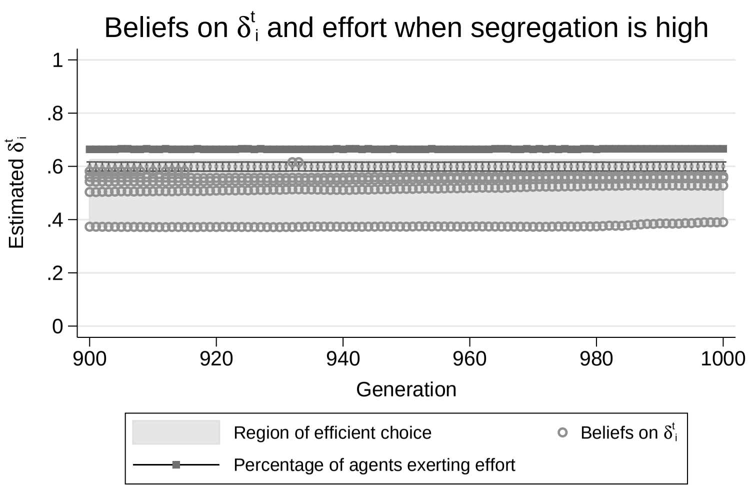
Conversely, if segregation is high or very high, beliefs do not converge towards a unique value. As a matter of fact, when segregation is high or very high, two distinct groups emerge. One group overestimates the returns to effort, while a second group is underestimating the returns to effort. That is, when the level of segregation is high or very high agents holding extreme beliefs are always exerting (no) effort. More generally, what this simulations show is that naivete and high levels of social segregation are sufficient conditions for the existence of persistent differences in beliefs.



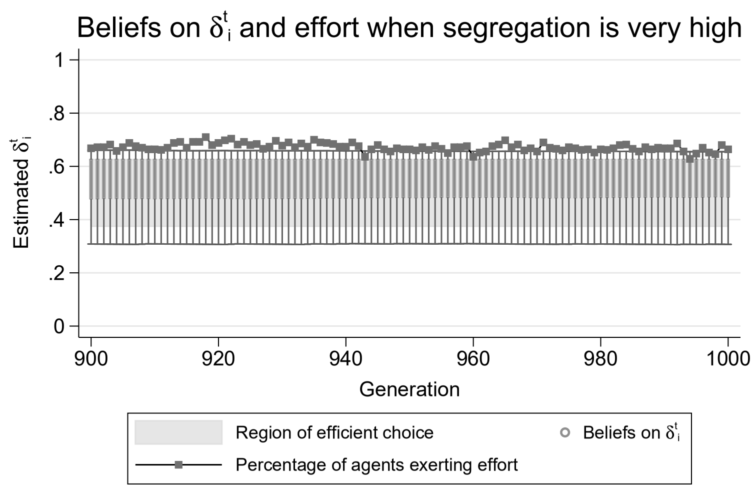
(A) $\psi = .8$



(B) $\psi = .6$



(C) $\psi = .4$



(D) $\psi = .2$

FIGURE C.1
Evolution of beliefs regarding δ.