# Sorting in the Presence of Misperceptions with an Application to Income Inequality*

Lisa Windsteiger†

November 8, 2016

**Abstract**

In this paper I analyze how social segregation and beliefs interact. Sorting decisions will be affected by beliefs about society, but these beliefs about society are in turn influenced by social interactions. In my model, people sort into social groups according to income, but become biased about the income distribution once they interact only with their own social circle. I examine, which types of misperceptions guarantee the existence of "biased sorting equilibria", i.e. stable partitions in which people want to stay in their chosen group, despite their acquired misperceptions about the other groups. I introduce a criterion – the consistency requirement – and characterize environments for which it selects a unique biased sorting equilibrium out of all possible stable partitions. The second part of the paper displays one possible application of this new framework to inequality and the demand for redistribution. I show that under segregation an increase in inequality can lead to a decline in perceived inequality and therefore to a fall in people's support for redistribution. I motivate my main assumptions with empirical evidence from a small survey that I conducted via Amazon Mechanical Turk.

**JEL Classification:** D31, D83, H24, H53
*Keywords:* Personal Income Distribution, Inequality, Redistribution

---

*I thank Ronny Razin, Stephane Wolton, Thomas Piketty, Mike Savage, Matt Levy, Daniel Laurison, Dominik Hangartner, Stephan Maurer, Frank Cowell and the participants of the LSE Microeconomic Theory Work in Progress Seminar for their helpful suggestions and comments.

†LSE Economics Department and STICERD, Email: L.V.Windsteiger@lse.ac.uk

# 1 Introduction

Who we choose to socialize with is often determined by our beliefs about others, about their qualities and their characteristics, and by our surmise about what effects their good and bad traits will have on ourselves. It is natural that we try to interact frequently with people who we think we can benefit from, be it in a material sense or simply because we enjoy their company.

On the other hand, our beliefs about society are likely to be influenced by our social interactions, by what and who we observe on a day-to-day basis. Depending on how diverse our social circles are, we might end up knowing a lot or very little about certain groups in society. Specifically, if we do not interact with some social groups, we are prone to develop distorted beliefs about what people are like in that group, about their characteristics and their traits. This in turn might influence who we choose to interact with in the first place and hence solidify and reinforce our attitudes and beliefs.

Take for example the question of how contact with ethnic minorities affects people's attitudes towards minorities. Dustmann and Preston (2001) show that looking at the effect of living in an area with high ethnic diversity on attitudes towards minorities can give a misleading answer. The reason for this is that we can at least to some degree decide where we want to live, and therefore people might live in ethnically diverse areas because they have a favourable attitude towards minorities in the first place - there is a two-way interaction between location choice and people's beliefs that needs to be taken into account.

Another example is parents' school choice for their kids. There is considerable evidence (see e.g. a 2007 Center on Education Policy report using National Educational Longitudinal Study (NELS) data from 1988-2000)[1] that private schools and state schools yield relatively similar learning outcomes if we control for pupils' family background. Nevertheless, parents are willing to pay a lot to live in areas with supposedly "good" schools (especially in the US) or to send their kids to private schools. However, these seem to be mainly parents who were privately educated themselves. In fact, Evans and Tilley (2011) show that in the UK parents who went to private schools are five times more likely to send their kids to private schools than state-educated parents (controlling for income). It almost seems as if there is one group of society that has been privately educated, believes that private schools are better and keeps sending their kids to private schools and another group that is state educated, and doesn't find private schools worth paying for (despite being perfectly able to afford private schools for their kids). In fact, this situation is modelled in a recent paper by Levy and Razin (2016).

Moreover, if we want to examine how changes in the economy - like an increase in income inequality, a reform in the education system or a surge in immigration - affect social groups and the belief system in society, we have to bear in mind that it doesn't suffice to look at the direct effect that these changes have on segregation and beliefs. Where society will end up in the long

[1] On a related note, Abdulkadiroglu et al. (2014) show that high achieving peers and racial composition of schools have no effect on learning achievement of individual pupils.

run depends also on the mutual reinforcement and interdepencence of social segregation and beliefs.

The interaction of social segregation and beliefs about society is what I examine in this paper. I take the canonical model of sorting according to income as a starting point. In this model, all feasible partitions of the income distribution are monotone, i.e. social groups will be single intervals of the income distribution. It is important to note, though, that without further assumptions there is nothing that pins down the exact way in which society will segregate in this model, i.e. how the social groups will look like. By varying the menu of sorting fees accordingly, any monotone partition of the income distribution is feasible. Furthermore, while this model takes into account that our beliefs about society affect our social interactions, it doesn't allow for the effect to go in the other direction: people's beliefs about the whole of society remain unchanged (and unbiased), even if people interact mainly with their own social circle.

One way to deal with the issue of multiplicity of equilibrium partitions in the canonical model is to explicitly model the supply side: We could assume that the sorting technology is offered by a profit-maximizing monopolist, by two or more competing firms or by a benevolent social planner and see whether this additional assumption pins down the menu of sorting fees and therefore the resulting equilibrium partition. The monopolist case has been examined by Rayo (2013), while Damiano and Li (2007) analyze the case of two or more competing firms and Levy and Razin (2015) examine the case of the social planner. In Windsteiger (2016), I compare the optimal partitions for the monopolist and the social planner and find that the optimal type of sorting depends on the shape of the income distribution and varies depending on which entity (profit-maximizing monopolist or benevolent social planner) is assumed to offer the sorting technology.

In the present paper I pursue a different path: I add misperceptions to the model. I demonstrate that this addresses two issues of the canonical model at the same time. First, it limits the amount of partitions that are feasible in equilibrium and therefore reduces multiplicity, and second it accounts for the fact that our social interactions shape our beliefs about society.

In my model of sorting with misperceptions I assume that, once society is segregated and people interact mainly with their own social circle, they become biased about the overall income distribution, and specifically about average income in the other groups. I define as "biased sorting equilibria" those partitions of the income distribution that are stable given people's misperceptions, i.e. partitions in which people want to stay in their chosen group, despite their acquired distorted beliefs. I show that adding misperceptions to the model initially leads to more complications: While in the canonical model all equilibrium partitions are monotone, i.e. single, connected intervals, biased sorting equilibria can also be non-monotone and hence people in one and the same group can have very different incomes, which complicates the analysis and is at odds with empirical evidence of assortative matching. Furthermore, the issue of multiplicity of equilibrium partitions persists, and even for a given equilibrium partition the sorting fees might not be uniquely determined. Finally, people's beliefs about

other groups can be inconsistent with what they see: they can be surprised by seeing people with certain incomes choosing to be in certain groups, because given their beliefs about incomes in the other groups they do not think these choices are optimal.

In order to address these problems I introduce a refinement criterion that I call the "consistency requirement". A partition satisfies consistency if the misperceptions are such that people are not surprised by the choices of people in other groups. I show that all biased sorting equilibria with consistency are monotone, and that the menu of sorting fees is uniquely pinned down for a given equilibrium partition. Furthermore, I demonstrate that if the misperceptions satisfy a form of monotonicity, then the consistency requirement selects a unique biased sorting equilibrium out of all possible stable partitions. In that case, social groups and sorting fees are uniquely determined.

The second part of the paper displays one possible application of this general framework to the issue of income inequality and the demand for redistribution. In this application, I assume that people's misperceptions are of a specific form: People overestimate the similarity (in terms of income) of the other groups, who they do not interact with, to their own group. This means that rich people overestimate poor people's income and poor people underestimate how much the rich earn. As a result, everybody in society underestimates income inequality. There is ample evidence that people are biased in this way, as work by Kiatpongsan and Norton (2014), Norton and Ariely (2011), Cruces et al. (2013) and also my own survey (see below) demonstrates. I therefore think it is of interest to investigate what effects biases of this form can have on social segregation and people's beliefs in my model.

I show that with these biases, segregation leads poor people to underestimate what they can gain from redistribution and therefore to show less support for redistribution than if they would have perfect knowledge of the income distribution. Moreover, an increase in inequality (in the form of a mean-preserving spread of the income distribution) always leads to a smaller increase in perceived inequality and therefore in the demand for redistribution than if people were unbiased. The reason for this is that people with income below average fully observe the fall of low incomes, but do not fully see the offsetting increase of high incomes. Therefore, they think that average income has decreased. But because their demand for redistribution is determined by the difference between their own income and (perceived) average income, and both decrease if people are biased, demand for redistribution increases less than if people are unbiased and know that average income hasn't changed. I show that the increase in inequality can even be such that perceived inequality declines and therefore people's support for redistribution falls.

In the last part of the paper I motivate my theoretical findings with suggestive empirical evidence from a small survey that I conducted via Amazon Mechanical Turk. I find evindence that people do indeed sort according to income to some extent. Moreover, people tend to have biased beliefs about the income distribution that confirm the assumptions that I make in the application of my model to the issue of redistributive preferences. Finally, I find that the

severity of those biases is increasing in the degree of social segregation of the respondents.

The rest of this paper is organized as follows. Section 2 discusses related literature, Section 3 presents the canonical model of sorting with respect to income and Section 4 introduces misperceptions into that model and explains the concept of biased sorting equilibrium and the consistency requirement. Section 5 finds conditions on the functional form of the misperceptions that lead to existence and uniqueness of biased sorting equilibria with consistency. Section 6 analyzes a two-group model where people overestimate the similarity (in terms of income) of their own group to the other group and applies this model to the issue of inequality and the demand for redistribution and Section 7 concludes.

# 2 Relation to existing literature

The canonical model of sorting and assortative matching was most famously employed by Becker (1974) to model the marriage market. Pesendorfer (1995) uses it to explain fashion cycles. Rayo (2013) examines optimal sorting from a profit-maximizing monopolist's point of view, while Levy and Razin (2015) analyze the welfare aspect of sorting and voters' preferences for redistribution in the presence of sorting. What all these papers have in common is that beliefs about society determine who people interact with, but social interactions do not influence beliefs. In fact, people retain perfect knowledge about society despite interacting only with a (potentially small) group of society.

Recently, the fact that segregation can affect beliefs has gained attention in the literature: On the theoretical side, Golub and Jackson (2012) present a model in which homophily (and resulting segregation) slows down convergence to a consensus in society. Concerning empirical evidence, Algan et al. (2015) show that political views converge among peers at university, and Boisjoly et al. (2006) and Burns et al. (2013) find that having roommates of a different ethnicity to one's own lowers students' prejudices.

That (potentially biased) beliefs can, in turn, have an effect on segregation, is pointed out by Dustmann and Preston (2001). They argue that estimating the effect of living in ethnically diverse neighbourhoods on attitudes towards minorities can lead to biased results, if we do not take into account how those attitudes affect neighbourhood choices in the first place. Levy and Razin (2016) present a model in which beliefs about school quality and parent's school choice for their children interact to create essentially two groups of society: a group of privately educated parents who believe in the benefits of the private school system and send their children to private school as well, and a group of state educated parents, who think private schools are not worth paying for and send their kids to state schools.

The first main contribution of my paper is to present a general model in which beliefs about society and segregation choice interact to create an endogenous system of beliefs and societal groups. This general model can be used to analyze sorting according to many variables that are distributed continuously in

society.[2] While in my version of the model, I assume that people sort according to income, the continuous variable could also be "ability" or "intelligence" and the model could be about sorting into different types of schools.

In the present paper I apply this general model to the situation of sorting according to income and preferences for redistribution. Standard political economy models (see e.g. Meltzer and Richard (1981)) predict that the demand for redistribution should be higher, the poorer the median earner is relative to average income in society. However, studies comparing pre-tax income inequality to redistribution rates in democracies, and hence trying to confirm the Meltzer-Richard Model empirically, deliver mixed results. Some papers do indeed find a positive link between inequality and redistribution (see e.g. Borge and Rattsoe (2004), Meltzer and Richard (1983) and Milanovic (2000)). However, others detect a negative relationship (e.g. Georgiadis and Manning (2012) and Rodriguez (1999)) or no significant link at all (e.g. Kenworthy and McCall (2008) and Scervini (2012)).

There are, of course, many explanations for why a high degree of inequality might not be reflected in high realized redistribution rates in an economy: Bartels (2009) argues that the views of the majority might be disregarded by political leaders due to successful lobbying of the financially powerful. Moreover, poor people might participate in the political process to a lesser degree than rich people, which might shift the identity of the median - decisive - voter (see e.g. Larcinese (2005)). Finally, and importantly, people rarely get to vote directly on redistribution rates. Instead, policital candidates offer platforms that take a position on a variety of issues, and people might vote against their interest on the subject of redistribution if they consider other issues to be more important (see Matakos and Xefteris (2016)).[3] However, apart from these factors, which affect *realized* redistribution rates, it seems to be the case that even the pure redistributive preferences of the population are not in line with what we might call the "Meltzer-Richard-Hypothesis": that pre-tax inequality and the demand for redistribution should be positively correlated, both across countries and over time (see e.g. Ashok et al. (2015)).

There are many ways to try and explain why high degrees of inequality do not necessarily correspond to high demand for redistribution in a country. In the Meltzer-Richard Model, people aim to maximize their own after-tax income and hence their sole concern is their relative position in the income distribution as a direct predictor of how much they would benefit or lose from redistribution. More sophisticated models allow for people's preferences for redistribution to be influenced also by other factors, such as social mobility or the overall degree of inequality in society (see e.g. Piketty (1995), Benabou and Ok (2001) and Alesina and Angeletos (2005)). This can explain why the median voter's relative position in the income distribution is not necessarily a good predictor of a society's demand for redistribution. However, also in these more elaborate models it will be the case that if inequality increases (ceteris paribus), demand

[2] Due to strict increasingness of $U$ those should, however, be variables where the whole of society agrees that "more is better", such as intelligence, ability, income or wealth.

[3] For a concise overview see Bonica et al. (2013).

for redistribution increases. Nevertheless, empirically we find that periods of increasing inequality can be accompanied by stagnant or declining demand for redistribution.

The second main contribution of my paper is that I show how my model of endogenous segregation and belief formation can be used to explain low support for redistribution in societies where inequality is high: As people interact only with people who have similar income to their own, they misperceive the shape of the whole income distribution, and poor people (including the median voter) underestimate how much they could gain from redistribution. Moreover, I demonstrate that with endogenous segregation and beliefs, the relationship between redistributive demand and inequality can be non-monotone - an increase in inequality can lead to a decline in the demand for redistribution, because people, if they see only a select group of society, might perceive that inequality has gone down due to the change in the income distribution. To my knowledge, there is no other theoretical model that finds these dynamics.

There is an growing empirical literature on people's misperceptions of the income distribution. Cruces et al. (2013) find that poor people overestimate their relative position in the income distribution, while rich people underestimate it. They also show that this lowers poor people's demand for redistribution: when their biases are corrected, poor people's demand for redistribution increases. Importantly, they additionally show that (social resp. economic) segregation affects people's misperceptions. Karadja et al. (2015) conduct a similar study for Sweden and find that a majority of people there tend to underestimate their relative position.

# 3 A theoretical model of economic segregation

In the following paragraphs, I will introduce the canonical model of sorting according to income that I use as a starting point for my own model of sorting with misperceptions. For reasons of expositional simplicity, I restrict my analysis in the main part of this paper to models with (a maximum of) two groups in society. In the supplementary Online Appendix, I show how the canonical model and the model of sorting with misperceptions can be extended to arbitrarily many groups.

Let income $y$ in an economy be distributed according to some income distribution $F(y)$, on the interval $Y = [0, y_{\max}]$ where $y_{\max} < \infty$. Assume furthermore that $F(y)$ is continuous and strictly monotonic. (As $F(y)$ is an income distribution, I will also assume that $F(y)$ is positively skewed (meaning that the median income is smaller than the average income), but this matters only for the application in Section 6.3 of the paper and is not necessary for the analysis of the general model.)

Suppose that an agent's utility is increasing not only in her own income but also in the average income of the people that she interacts with, which I will henceforth call her "reference group". Specifically, I will assume that a person with income $y_j$ gets utility $U_j = U(y_j, E(y|y \in S_i))$, where $S_i$ is individual $j$'s

reference group. If there is no economic segregation, let everybody's reference group be a representative sample of the whole population, such that $U_j = U(y_j, E(y))$. However, suppose that a person with income $y_j$ can pay $b$ to join group $S_b$ and get utility

$$U(y_j, E[y|y \in S_b]) - b$$

or refrain from paying $b$ and get

$$U(y_j, E[y|y \in S_0])$$

where $S_b$ is the set of incomes $y$ of people who have paid $b$ and $S_0$ is the set of incomes $y$ of people who haven't paid $b$. Let $U(.,.)$ be continuous, strictly increasing in both arguments and strictly supermodular, such that[4]

$$\forall x' > x : U(y, x') - U(y, x) \text{ is strictly increasing in } y$$

Then I can define the following:

**Definition 1** *A* ***sorting equilibrium*** *is a partition $\{\{S_0\}, \{S_b\}\}$ of $Y$ and a sorting fee $b > 0$ such that*

$$U(y, E[y|y \in S_b]) - b \leq U(y, E[y|y \in S_0]) \quad \forall y \in S_0 \tag{1}$$
$$U(y, E[y|y \in S_b]) - b \geq U(y, E[y|y \in S_0]) \quad \forall y \in S_b \tag{2}$$

In a sorting equilibrium as defined above people stay in the group that gives them the highest utility.

**Corollary 1** *In any sorting equilibrium, $S_b$ corresponds to higher average income than $S_0$.*

**Proof.** This immediately follows from Definition 1, from $b > 0$ and from the fact that $U$ is strictly increasing in both arguments. ■

[4] Note that this paper offers only a very reduced form model of economic segregation. That people's utility is increasing in the average income of the other people they mix with is perhaps a simplified way of saying that living in an affluent neighborhood offers many benefits, such as good schools (because people are willing to spend more on the education of their kids, and because the presence of children of rich people might increase other pupil's chances in life through various peer effects) and pleasant surroundings such as parks or leisure centres (perhaps with increased security or surveillance). Instead of modelling all this on a micro level, I subsum all these effects into a utility function which is increasing in the average income of one's peers.

I can show that all sorting equilibria will be of a certain form:

**Proposition 1** *All sorting equilibria will be monotone.*[5]

**Proof.** Suppose w.l.o.g. that a sorting equilibrium exists where $y_1 \in S_b$ and $y_2 \in S_0$, with $y_2 > y_1$. Then I must have

$$U(y_2, E[y|y \in S_b]) - U(y_2, E[y|y \in S_0]) \leq b$$

and

$$U(y_1, E[y|y \in S_b]) - U(y_1, E[y|y \in S_0]) \geq b$$

and hence

$$U(y_1, E[y|y \in S_b]) - U(y_1, E[y|y \in S_0]) \geq U(y_2, E[y|y \in S_b]) - U(y_2, E[y|y \in S_0])$$

But due to $y_2 > y_1$ this is a contradiction to $U$ being strictly supermodular. ■

Proposition 1 allows me to rewrite the definition of a sorting equilibrium in terms of an equilibrium cutoff $\hat{y}$. [6]

**Corollary 2** *A* ***sorting equilibrium*** *is a partition of $Y$, characterized by an equilibrium cutoff $\hat{y}$, and a sorting fee $b > 0$ such that*

$$U(\hat{y}, E[y|y \geq \hat{y}]) - U(\hat{y}, E[y|y < \hat{y}]) = b \tag{3}$$

**Proof.** Given the fact that $S_0 = [0, \hat{y})$ and $S_b = [\hat{y}, y_{\max}]$ and the equilibrium condition from Definition 1, it follows that both

$$U(y, E[y|y \geq \hat{y}]) - U(y, E[y|y < \hat{y}]) < b \quad \forall y \in [0, \hat{y})$$

and

$$U(y, E[y|y \geq \hat{y}]) - U(y, E[y|y < \hat{y}]) \geq b \quad \forall y \in [\hat{y}, y_{\max}]$$

need to hold in any sorting equilibrium.[7] Due to supermodularity of $U$, these two conditions can be simplified to

$$U(\hat{y}, E[y|y \geq \hat{y}]) - U(\hat{y}, E[y|y < \hat{y}]) \leq b$$

[5] By *monotone* I mean that the groups $S_0$ and $S_b$ are single intervals of $Y$, i.e. they are of the form $S_0 = [0, \hat{y})$ and $S_b = [\hat{y}, y_{\max}]$ for some cutoff $\hat{y}$. (By Corollary 1, $S_b$ has to be to the right of $S_0$ on the $Y$ scale.)

[6] I will for the rest of the paper choose the convention to define the partition characterized by $\hat{y}$ such that the person at $\hat{y}$ (who in equilibrium is indifferent between being in $S_0$ and $S_b$) belongs to the rich group, i.e. $S_0 = [0, \hat{y})$ and $S_b = [\hat{y}, y_{\max}]$.

[7] Equilibrium condition (1) translates to

$$U(y, E[y|y \geq \hat{y}]) - U(y, E[y|y < \hat{y}]) \leq b \quad \forall y \in [0, \hat{y})$$

but strict supermodularity of $U$ implies that in equilibrium nobody in $[0, \hat{y})$ can be indifferent between being in $S_0$ and $S_b$ because then the person who is just $\varepsilon$ to the right of her (and therefore also in $[0, \hat{y})$) would be better off in $S_b$. Hence strict inequality must hold on $[0, \hat{y})$.

and

$$U(\hat{y}, E[y|y \geq \hat{y}]) - U(\hat{y}, E[y|y < \hat{y}]) \geq b$$

Hence, we get

$$U(\hat{y}, E[y|y \geq \hat{y}]) - U(\hat{y}, E[y|y < \hat{y}]) = b.$$

This implies that a person with income $\hat{y}$ just at the border of the two groups has to be exactly indifferent between joining either of the two groups in equilibrium. ■

**Corollary 3** *For any equilibrium partition, the corresponding sorting fee $b > 0$ will be unique.*

**Proof.** This follows immediately from equilibrium condition (3). ■

The above presented canonical model of sorting according to income has many positive features: The equilibrium partitions will always be monotone and therefore individual income within the resulting equilibrium groups will be similar (or at least within a simple interval), which simplifies the analysis of the model and is also compatible with empirical evidence of segregation according to income and assortative matching. Furthermore, inherent in the above definition is a notion of *consistency*: if a person sees the income of another person and knows which group this person joined, she always thinks that this person is correct in doing so, because both people evaluate the benefit of being in a certain group (given a certain income) equally. This means that no extra condition is needed to guarantee consistency, it "comes for free" in the equilibrium condition (3) - if a person in a certain group thinks that being in that group is best for her, then also all other people - no matter which group they belong to - will think that this is optimal for her.

However, from equilibrium condition (3) it is immediate to see that the model delivers no prediction about the type of segregation that will happen in a society, i.e. how the social groups will look like: For any continuous distribution function $F$ there exists an infinite number of equilibrium partitions. More specifically, for any cutoff $\hat{y} \in Y$ there exists a sorting fee $b$ such that $(\hat{y}, b)$ is a sorting equilibrium. Moreover, while in this model people's beliefs about society determine their social interactions, the reverse effect is not taken into account: segregation has no effect on people's beliefs about the economy and people retain perfect knowledge about the income distribution in the whole of society, even though they interact mainly with a select group of people who are similar to them in terms of income.

One way to try and resolve the issue of multiplicity is to look at the supply side of the sorting technology: we can analyze the optimal partition that a profit-maximizing monopolist, a number of competing firms or a benevolent social planner would want to offer and thereby select "plausible" equilibria out of the infinite number of possible equilibria. I explore this path in another paper of mine and find that the form of resulting optimal partitions depends on the underlying distribution function and on which entity is assumed to provide the sorting technology (see Windsteiger (2016)).

In this paper, I pursue a different path: I add misperceptions to the model. Specifically, I will assume that people, once they are sorted into their respective groups, become biased about average income in the other groups, and I will define partitions as biased sorting equilibria if they are such that people want to stay in their group given their misperceptions about the rest of society. I will show in the next section that this addresses both of the above mentioned limitations of the canonical model: First, it lifts the assumption that people retain perfect knowledge about society once they are sorted and therefore allows for the interaction between segregation and beliefs to go both ways. Second, restricting attention to biased sorting equilibria with the additional requirement of consistency (which I will explain below) greatly reduces the number of possible equilibrium partitions and can lead, if the misperceptions are of a certain form, even to uniqueness.

# 4 Sorting with misperceptions

In the following analysis, I will restrict my attention to monotone partitions, i.e. partitions $P = [S_0, S_b]$ of $Y$ that can be uniquely characterized by a cutoff $\hat{y} \in Y$ (with the convention that $S_0 = [0, \hat{y})$ and $S_b = [\hat{y}, y_{\max}]$), and I will henceforth call the people in $S_0$ "the poor" and the people in $S_b$ "the rich". Without further assumptions also non-monotone equilibria are possible if people have misperceptions. In Appendix A.1 I show that restricting the analysis to monotone partitions is without loss of generality for the analysis that I conduct in this paper.[8]

Suppose that people, once they are sorted into their group, become biased about average income in the other group and hence about the overall income distribution. I will model a group's belief about the other group as resulting from a group belief "technology". Specifically, I will assume that people's biased perception of the average income of the other group can be characterized by the continuous belief function

$$B : \mathbf{P} \to Y^4$$

where $\mathbf{P}$ is the space of all monotone partitions $P = [S_0, S_b]$ of $Y$. $B$ is a continuous function that maps all monotone partitions of $Y$ (and note that any monotone partition can be uniquely characterized by the cutoff $\hat{y}$) into a four-dimensional vector of beliefs

$$B(\hat{y}) = (\underline{E}(\hat{y}), \bar{E}_p(\hat{y}), \underline{E}_r(\hat{y}), \bar{E}(\hat{y}))$$

where the first two entries denote the poor group's belief about average income in the poor and the rich group respectively and the last two entries denote the rich group's belief about average income in the poor and the rich group. $\underline{E}(\hat{y})$ is the true average income in the poor group, i.e. $\underline{E}(\hat{y}) = E[y|y < \hat{y}]$ and $\bar{E}(\hat{y})$ is the correct average income in the rich group, $\bar{E}(\hat{y}) = E[y|y \geq \hat{y}]$. The poor's

[8] I show that the refinement that I introduce in this section (the consistency requirement), implies monotonicity.

belief about average income in the rich group is $\bar{E}_p(\hat{y})$ and the rich's belief about average income in the poor group is $\underline{E}_r(\hat{y})$.

This definition of the belief function implies that people are correct about average income in their own group. Furthermore, misperceptions are constant within groups, i.e. people who are in the same group have the *same* misperception about the other group's average (and thus misperceptions do not depend on one's own income directly, but on group membership).[9]

I also restrict the beliefs of one group about average income in the other group to actually lie in that group's income range:

**Assumption 1** *The belief function $B(\hat{y})$ is such that*

$$\bar{E}_p(\hat{y}) \in [\hat{y}, y_{\max}] \text{ and } \underline{E}_r(\hat{y}) \in [0, \hat{y}] \ \forall \hat{y} \in Y$$

Given this belief function, I can define the following:

**Definition 2** *A monotone partition of $Y$ (characterized by an equilibrium cutoff $\hat{y}^*$) and a sorting fee $b > 0$ constitute a* ***biased sorting equilibrium*** *iff*

$$\begin{aligned} U(y, \bar{E}_p(\hat{y}^*)) - b &\leq U(y, \underline{E}(\hat{y}^*)) \quad \forall y \in [0, \hat{y}^*) && \text{(IC1)} \\ U(y, \bar{E}(\hat{y}^*)) - b &\geq U(y, \underline{E}_r(\hat{y}^*)) \quad \forall y \in [\hat{y}^*, y_{\max}] && \text{(IC2)} \end{aligned}$$

A biased sorting equilibrium is therefore a partition of $Y$ that is "stable" given people's misperceptions about the other group. When people compare the utility they obtain in their own group to the utility they *think* they could obtain in the other group - given their misperceptions about average income in the other group - they come to the conclusion that they reach the highest possible level of utility in their own group and therefore they do not want to move to the other group.

**Corollary 4** *The equilibrium sorting fee $b$ is not necessarily unique in a biased sorting equilibrium.*

**Proof.** First, note again that due to strict supermodularity of $U$, inequality (IC1) has to be strict in equilibrium for all $y$ in $[0, \hat{y}^*)$. Therefore, (IC1) and (IC2) combined imply

$$U(\hat{y}^*, \bar{E}(\hat{y}^*)) - U(\hat{y}^*, \underline{E}_r(\hat{y}^*)) \geq b \geq U(\hat{y}^*, \bar{E}_p(\hat{y}^*)) - U(\hat{y}^*, \underline{E}(\hat{y}^*))$$

If the beliefs are such that one of the inequalities is strict, there exists a range of $b$ such that this condition is satisfied. ■

[9] I chose to restrict my assumption to misperceptions that are constant within group because I specifically want to focus on differences in perceptions *between* groups rather than *within* groups. This restriction helps to simplify the analysis, but the main results of this paper would not change fundamentally if biases were to vary also within groups. The restriction can be deduced "naturally" from the assumption that people interact and communicate freely within their own group and hence will, within their group, reach a common belief about the other groups.

## 4.1 Consistency

At first, adding misperceptions to the model does not simplify the analysis, but rather adds some additional problems: The problem of multiplicity of equilibria is not solved, on the contrary, depending on the belief function, even non-monotone partitions can be biased sorting equilibria (see Appendix A.1). Furthermore, as Corollary 4 shows, the sorting fee $b$ need no longer be uniquely determined. Moreover, we do not necessarily have the notion of consistency (which is inherent in the unbiased model, as explained above) in a biased sorting equilibrium. It can be that people's beliefs about the other group are inconsistent with what they see: A person in the poor group might wonder why a person in the rich group finds it worthwhile to pay $b$ given the poor person's belief about average income in the rich group. Similarly, a person in the rich group might - given the rich group's misperception about average income in the poor group - wonder why a certain person in the poor group doesn't want to join the rich group.

However, the inconsistency and the non-uniqueness of the menu of sorting fees for a given equilibrium partition vanish if I introduce what I call the *consistency requirement*. It requires that people who are in the poor group think that the people who are in the rich group are correct in doing so and vice versa. Formally, this requirement translates to

**Definition 3** *A monotone partition of $Y$ characterized by a cutoff $\hat{y}$ and a sorting fee $b$ satisfy **consistency** iff*

$$U(y, \bar{E}(\hat{y})) - b \leq U(y, \underline{E}_r(\hat{y})) \quad \forall y \in [0, \hat{y}) \qquad \text{(CR1)}$$
$$U(y, \bar{E}_p(\hat{y})) - b \geq U(y, \underline{E}(\hat{y})) \quad \forall y \in [\hat{y}, y_{\max}] \qquad \text{(CR2)}$$

In words, condition (CR1) requires that a person in the rich group who looks at any person with income $y$ in the poor group thinks that this person cannot achieve higher utility by switching to the rich group (and note that the person from the rich group evaluates person $y$'s utility in the poor group given her own biased perception of average income in the poor group, $\underline{E}_r(\hat{y})$). Condition (CR2) does the same for poor people's belief about the rich group.

As I have pointed out, in the "unbiased" sorting equilibrium that I have defined in the previous section, consistency is implicit. Because everybody has the same (correct) understanding of average incomes in both groups, people cannot be "puzzled" by other people's choices - everybody evaluates everybody else's utility in the same way. It is only when people have incorrect perceptions of the other group that consistency becomes a separate issue and is not implicit in the equilibrium definition. People can be happy with their own choices (which means the partition constitutes a biased sorting equilibrium), while at the same time not understanding other people's choices (which means that consistency is violated). Hence, it makes sense - as a refinement to biased sorting equilibria - to define biased sorting equilibria which additionally satisfy consistency:

**Definition 4** *A monotone partition of $Y$ (characterized by an equilibrium cutoff $\hat{y}^*$) and a sorting fee $b > 0$ constitute a* ***biased sorting equilibrium with consistency*** *iff*

$$\begin{aligned} U(y, \bar{E}_p(\hat{y}^*)) - b &\leq U(y, \underline{E}(\hat{y}^*)) \quad \forall y \in [0, \hat{y}^*) && \textit{(IC1)} \\ U(y, \bar{E}(\hat{y}^*)) - b &\geq U(y, \underline{E}_r(\hat{y}^*)) \quad \forall y \in [\hat{y}^*, y_{\max}] && \textit{(IC2)} \end{aligned}$$

*and*

$$\begin{aligned} U(y, \bar{E}(\hat{y}^*)) - b &\leq U(y, \underline{E}_r(\hat{y}^*)) \quad \forall y \in [0, \hat{y}^*) && \textit{(CR1)} \\ U(y, \bar{E}_p(\hat{y}^*)) - b &\geq U(y, \underline{E}(\hat{y}^*)) \quad \forall y \in [\hat{y}^*, y_{\max}] && \textit{(CR2)} \end{aligned}$$

**Corollary 5** *A monotone partition of $Y$ (characterized by a cutoff $\hat{y}^*$) and a sorting fee $b > 0$ constitute a biased sorting equilibrium with consistency iff*

$$\begin{aligned} & U(\hat{y}^*, \bar{E}_p(\hat{y}^*)) - U(\hat{y}^*, \underline{E}(\hat{y}^*)) \\ &= U(\hat{y}^*, \bar{E}(\hat{y}^*)) - U(\hat{y}^*, \underline{E}_r(\hat{y}^*)) \\ &= b \end{aligned} \tag{4}$$

**Proof.** (IC1) and (CR2) can be combined to[10]

$$U(\hat{y}^*, \bar{E}_p(\hat{y}^*)) - U(\hat{y}^*, \underline{E}(\hat{y}^*)) = b \tag{5}$$

Analogously, (IC2) and (CR1) can be combined to

$$U(\hat{y}^*, \bar{E}(\hat{y}^*)) - U(\hat{y}^*, \underline{E}_r(\hat{y}^*)) = b \tag{6}$$

Together, (5) and (6) give the equilibrium condition (4). ■

Hence, a biased sorting equilibrium with consistency is a partition where the perceived benefit of being in the rich group rather than the poor group (in terms of utility) of the person with income at the equilibrium cutoff $\hat{y}^*$ is regarded to be equally high by both groups. Note that for a given equilibrium cutoff $\hat{y}^*$ that satisfies (4), the corresponding sorting fee $b$ is unique.

The equilibrium condition (4) restricts the set of belief functions which imply equilibrium existence. In the next sections, I derive conditions on this function such that equilibrium exists and is unique.

# 5 Different types of misperceptions

What kind of biased sorting equilibria can exist for different types of misperceptions? In the following analysis, I restrict my attention to misperceptions where the direction of the bias does not vary with the cutoff, i.e. for any configuration of groups one group either always overestimates or underestimates

[10] Again, note that inequalities (IC1) and (CR1) need to be strict in any equilibrium $\forall y \in [0, \hat{y}^*)$.

average income in the other group and groups do not switch between over- and underestimating each other depending on group size or shape. Formally, this means I look only at belief functions $B$ such that

$$\bar{E}_p(\hat{y}) < (>) \bar{E}(\hat{y}) \quad \forall \hat{y} \in (0, y_{\max})$$

and

$$\underline{E}_r(\hat{y}) > (<) \underline{E}(\hat{y}) \quad \forall \hat{y} \in (0, y_{\max}),$$

Note, though, that by Assumption 1 a group is correct about average income in the other group if they see the whole income distribution, i.e. the poor are correct at $y_{\max}$,

$$\bar{E}_p(\hat{y}) = \bar{E}(\hat{y}) \quad \text{if} \quad \hat{y} = y_{\max},$$

and the rich are correct at 0,

$$\underline{E}_r(\hat{y}) = \underline{E}(\hat{y}) \quad \text{if} \quad \hat{y} = 0$$

Then there are four possible combinations of biases:

Case 1: $\underline{E}_r(\hat{y}) > \underline{E}(\hat{y})$ and $\bar{E}_p(\hat{y}) < \bar{E}(\hat{y}) \quad \forall \hat{y} \in (0, y_{\max})$: The rich overestimate average income in the poor group and the poor underestimate average income in the rich group.

Case 2: $\underline{E}_r(\hat{y}) < \underline{E}(\hat{y})$ and $\bar{E}_p(\hat{y}) > \bar{E}(\hat{y}) \quad \forall \hat{y} \in (0, y_{\max})$: The rich underestimate average income in the poor group and the poor overestimate average income in the rich group.

Case 3: $\underline{E}_r(\hat{y}) < \underline{E}(\hat{y})$ and $\bar{E}_p(\hat{y}) < \bar{E}(\hat{y}) \quad \forall \hat{y} \in (0, y_{\max})$: Both groups underestimate each other's average income

Case 4: $\underline{E}_r(\hat{y}) > \underline{E}(\hat{y})$ and $\bar{E}_p(\hat{y}) > \bar{E}(\hat{y}) \quad \forall \hat{y} \in (0, y_{\max})$: Both groups overestimate each other's income.

In the first case, both groups underestimate the difference between groups, while in the second case they both overestimate it. In the latter two cases the misperceptions work in opposite directions for the two groups: one group overestimates the difference, the other group underestimates it. In Appendix A.2 I analyze these four combinations to see whether biased sorting equilibria (with and without consistency) can exist and I provide sufficient conditions for existence and uniqueness. It turns out that interior biased sorting equilibria with consistency can only exist in two of the four possible combinations: either both groups think the other group is more similar to themselves or both groups think the other group is more different to themselves. The reason for this is that the equilibrium condition (4) requires both groups to have the same perception of the benefits of sorting at the equilibrium cutoff, and hence it cannot be the case that one group underestimates the difference between the groups for any cutoff, while the other group overestimates it.

**Proposition 2** *No cutoff $\hat{y} \in Y$ constitutes a biased sorting equilibrium in Case 4.*

**Proposition 3** *Any cutoff $\hat{y} \in Y$ constitutes a biased sorting equilibrium in Case 3, but no cutoff $\hat{y} \in Y$ constitutes a biased sorting equilibrium with consistency.*

**Proof.** See Appendix A.2. ■

While biased sorting equilibria with consistency cannot exist in Case 3 and 4, it turns out that they will always exist in Case 1 and 2.

**Proposition 4** *There always exists a biased sorting equilibrium with consistency in Case 1 and 2.*

**Proof.** See Appendix A.2. ■

The intuition for why Case 1- and Case 2-type misperceptions guarantee the existence of equilibria is that the form of the misperceptions ensures that the groups' perceived benefits of sorting cross at least once: At $\hat{y} = 0$ the poor's perceived benefits of sorting are lower (higher) than the rich's, and the reverse is true at $\hat{y} = y_{\max}$. As all functions are continuous, there must be an interior cutoff $\hat{y} \in (0, y_{\max})$ such that they are equal.

If $B$ and $U(.,.)$ are such that the perceived benefits of sorting of the two groups intersect only once, the interior equilibrium cutoff is unique. Sufficient conditions for this are stated in Appendix A.2.3. If $U$ is linear in both arguments[11], the sufficient conditions for uniqueness simplify to the following:

**Proposition 5** *(Case 1 Uniqueness) If $U(.,.)$ is linear, the biased sorting equilibrium with consistency is unique if the misperceptions monotonically converge to the truth, i.e.*

$$\frac{d(\bar{E}(\hat{y}) - \bar{E}_p(\hat{y}))}{d\hat{y}} < 0 \quad \textit{and} \quad \frac{d(\underline{E}_r(\hat{y}) - \underline{E}(\hat{y}))}{d\hat{y}} > 0 \quad \forall \hat{y} \in (0, y_{\max})$$

**Proof.** See Appendix A.2.1. ■

**Proposition 6** *(Case 2 Uniqueness) If $U(.,.)$ is linear, the interior equilibrium is unique if the misperceptions monotonically converge to the truth, i.e.*

$$\frac{d(\bar{E}_p(\hat{y}) - \bar{E}(\hat{y}))}{d\hat{y}} < 0 \quad \textit{and} \quad \frac{d(\underline{E}(\hat{y}) - \underline{E}_r(\hat{y}))}{d\hat{y}} > 0 \quad \forall \hat{y} \in (0, y_{\max})$$

**Proof.** See Appendix A.2.1. ■

The condition that mispeceptions converge to the truth monotonically as the cutoff goes to 0 resp. $\hat{y}$, i.e.

$$\frac{d\left|\bar{E}(\hat{y}) - \bar{E}_p(\hat{y})\right|}{d\hat{y}} < 0 \text{ and } \frac{d\left|\underline{E}_r(\hat{y}) - \underline{E}(\hat{y})\right|}{d\hat{y}} > 0$$

[11] If $U$ is linear then it is not strictly increasing whenever one of the arguments is 0. See Appendix A.2.1 for a specific analysis of linear utility functions at 0.

can be interpreted as people being less biased, the more they see of the income distribution: as $\hat{y}$ increases, the poor see a bigger part of the income distribution and their belief about average income in the other group becomes more accurate. The opposite happens for the rich: as the cutoff increases, they see a smaller part of society and therefore become more biased. In the limit this yields what I have already assumed in Assumption 1: the poor are correct at $y_{\max}$ and the rich at 0.

# 6 Underestimating Inequality

In the following section I will examine misperceptions of the Case 1 type for a specific belief function. In Case 1, the rich overestimate average income in the poor group and the poor underestimate average income in the rich group and hence both groups overestimate the similarity of the other group to their own group.

As I mention in the previous section, biased sorting equilibria with consistency exist in two cases: if both groups underestimate between-group differences (Case 1), and if both groups overestimate them (Case 2). The reason I decide to focus on Case 1 for the following sections is that I want to show how my model can be applied to the issue of income inequality and the demand for redistribution. There is ample evidence that, when thinking about the shape of the income distribution, people tend to have misperceptions that make them underestimate income inequality: I find evidence for biases in my own survey (see Section 6.6). Kiatpongsan and Norton (2014) examine perceptions of pay differences between CEOs and unskilled workers in 40 countries and find that people tend to vastly underestimate those differences. The fact that people underestimate inequality is well documented, see e.g. Norton and Ariely (2011) for evidence on the US wealth distribution and Norton et al. (2014) for Australia. I therefore think that it is of interest to analyze the model with misperceptions of the Case 1 type and examine what those misperceptions imply for people's demand for redistribution.[12]

Formally, I assume that the belief function $B$ is such that the people in the poor group think that the average income in the rich group is

$$\bar{E}_p(\hat{y}) = \beta(1 - F(\hat{y}))\hat{y} + (1 - \beta(1 - F(\hat{y}))\bar{E} \tag{7}$$

and the people in the rich group think that the average income of the poor is

$$\underline{E}_r(\hat{y}) = \gamma F(\hat{y})\hat{y} + (1 - \gamma F(\hat{y}))\underline{E} \tag{8}$$

$\beta \in [0, 1]$ and $\gamma \in [0, 1]$ parameterize the "naivity" of agents and if $\beta$ resp. $\gamma$ is 0 agents have no misperceptions. It is straightforward to see that $\bar{E}_p(\hat{y}) < \bar{E}(\hat{y})$ and $\underline{E}_r(\hat{y}) > \underline{E}(\hat{y})$ for all $y \in (0, y_{\max})$. The functional form of $\bar{E}_p(\hat{y})$ and $\underline{E}_r(\hat{y})$ implies that the misperceptions are more severe, the smaller the part of the

[12] I analyze Case-2-type misperceptions in Appendix A.10 and compare Case 1 and Case 2 in Appendices A.11 and A.12.

distribution that they can fully observe (which is $F(\hat{y})$ for the poor group and $1 - F(\hat{y})$ for the rich group). Specifically, we have that

$$\frac{d(\bar{E}(\hat{y}) - \bar{E}_p(\hat{y}))}{d\hat{y}} = -\beta(1 - F(\hat{y})) < 0 \quad \forall \hat{y} \in (0, y_{\max})$$

and

$$\frac{d(\underline{E}_r(\hat{y}) - \underline{E}(\hat{y}))}{d\hat{y}} = \gamma F(\hat{y}) > 0 \quad \forall \hat{y} \in (0, y_{\max})$$

therefore that the misperceptions converge to the truth monotonically as $\hat{y}$ goes to 0 resp. $y_{\max}$.[13]

Misperceptions of this type could arise in the following way: As people live in their segregated communities, they see mostly people who have income $y$ similar to their own (i.e. people from their own group). They do meet people from the other group, but they are not aware that most of the time they do not meet a representative sample of the other group (because they are more likely to meet people from the other group who are close to the cutoff). They see the average income in their own group, but what matters for their sorting decision is also the average income in the other group, which they do not see. Because they know $\hat{y}$ and the overall range of $y$ (i.e. that $y$ ranges from 0 to $y_{\max}$), they know that the average income of the other group lies somewhere between the cutoff $\hat{y}$ and 0 resp. $y_{\max}$. However, as they neglect the fact that they often do not meet a representative sample of the other group and are rather more likely to meet people very close to the cutoff, the poor think that the average in the rich group is closer to their own average than it actually is, and the same holds for the rich when thinking about the poor group's average. In short, people below the cutoff **underestimate** the average $y$ in the rich group and people above the cutoff **overestimate** the average $y$ in the poor group. This will lead both groups to underestimate the benefits of sorting: The rich because they think the poor are less poor than they actually are, and the poor because they think the rich are not as rich as they actually are.[14]

[13] For the following analysis it is not necessary that the misperceptions are of exactly of the form (7) and (8) . For the results of the next section to hold, I need the misperceptions to be such that a binary biased sorting equilibirum exists and is (ideally) unique. Sufficient conditions for this are stated in the previous section. Furthermore, I need the equilibrium cutoff to be located above median income. In Appendix A.2.4, I specify sufficient conditions on the belief function to guarantee that there is a unique interior equilibrium cutoff above the median.

[14] The specific form of misperception that I use in this paper can be microfounded in the following way: people in the poor group only sometimes encounter a representative sample of the rich (e.g. if they go to the opera, watch a royal wedding or shop in a fancy store) and the rest of the time encounter only rich people who are very close to the cutoff (basically at $\hat{y}$), maybe because they are parents of their kids' school friends (upper-middle class families sometimes prefer to send their kids to state schools). However, people are not aware of this and therefore estimate average income as if they were observing a representative sample of the other group. The particular functional form of the bias can arise if the frequency of meeting a representative sample of the other group depends on the size of the own group, $F(\hat{y})$. This can happen because the rich have less leeway to segregate if they are only a small group (perhaps it doesn't pay off to create a lot of extra posh golf clubs or wellness clubs if it is just the top 1% who would be members).

The functional form of the misperceptions as given by (7) and (8) is such that the sufficient conditions for existence of a biased sorting equilibrium with consistency are satisfied: $\bar{E}_p(\hat{y})$ and $\underline{E}_r(\hat{y})$ are continuous functions and each group is correct at one of the endpoints, whereas the other group is maximally biased at that respective endpoint. Furthermore, the misperceptions converge to the truth monotonically, and therefore there exists a unique interior equilibrium cutoff if the utility function is linear. For reasons of expositional simplicity, I will assume that $U(y, x) = yx$ for the following analysis, to avoid having to require all the extra assumptions that are needed to guarantee uniqueness with a general utility function (see Proposition 4).

The unique equilibrium cutoff with consistency can be calculated via the equilibrium condition

$$\hat{y}^*[\bar{E}(\hat{y}^*) - \underline{E}_r(\hat{y}^*)] = \hat{y}^*[\bar{E}_p(\hat{y}^*) - \underline{E}(\hat{y}^*)] \tag{9}$$

and note that the expressions on both sides also need to be equal to some $b > 0$, which rules out $\hat{y} = 0$ as an equilibrium cutoff (see also Appendix A.2.1). As the misperceptions satisfy the sufficient conditions for a unique equilibrium cutoff, we know that there is a unique $\hat{y}^* > 0$ that satisfies

$$\bar{E}(\hat{y}^*) - \underline{E}_r(\hat{y}^*) = \bar{E}_p(\hat{y}^*) - \underline{E}(\hat{y}^*) \tag{10}$$

Plugging in the functional form of the misperceptions, (7) and (8), and rearranging gives

$$\beta(1 - F(\hat{y}^*))(\bar{E}(\hat{y}^*) - \hat{y}^*) = \gamma F(\hat{y}^*)(\hat{y}^* - \underline{E}(\hat{y}^*))$$

and thus

$$\hat{y}^* = \frac{\beta(1 - F(\hat{y}^*))\bar{E}(\hat{y}^*) + \gamma F(\hat{y}^*)\underline{E}(\hat{y}^*)}{\beta(1 - F(\hat{y}^*)) + \gamma F(\hat{y}^*)}$$

which can be rewritten as

$$\hat{y}^* = \frac{a(1 - F(\hat{y}^*))\bar{E}(\hat{y}^*) + F(\hat{y}^*)\underline{E}(\hat{y}^*)}{a(1 - F(\hat{y}^*)) + F(\hat{y}^*)} \tag{11}$$

where $a = \beta/\gamma$.[15] Hence, $\hat{y}^*$ must be the fixed point of the function

$$h(\hat{y}) = \frac{a(1 - F(\hat{y}^*))\bar{E}(\hat{y}^*) + F(\hat{y}^*)\underline{E}(\hat{y}^*)}{a(1 - F(\hat{y}^*)) + F(\hat{y}^*)}$$

In Appendix A.2.5 I analyze the function $h(\hat{y})$ in detail (and I verify that it has indeed a unique fixed point $\hat{y}^*$). If $a = 1$ (and hence $\beta = \gamma$), (11) simplifies to $\hat{y}^* = E$ and the unique binary biased sorting equilibrium is such that the cutoff is exactly at the mean.

[15] $a > 0$ if both types are assumed to be naive to some degree, i.e. $\beta > 0$ and $\gamma > 0$. If one of the groups would be fully sophisticated, e.g. $\gamma = 0$, while the other group is naive, then consistency couldn't be satisfied for any (interior) cutoff. If both groups are fully sophisticated, we are back in the model of unbiased sorting of Section 3.

**Proposition 7** *If bias $\underline{E}_r$ and $\bar{E}_p$ are defined according to (7) and (8), there exists a unique interior biased sorting equilibrium with consistency, and the unique cutoff $\hat{y}^* > 0$ is the fixed point of $h(\hat{y}) = \frac{a(1-F)\bar{E}+F\underline{E}}{a(1-F)+F}$ where $a = \beta/\gamma$. If $a = 1$, the unique cutoff is at $\hat{y}^* = E$.*

## 6.1 The relationship between naivety and the cutoff

As noted above, the equilibrium cutoff depends on the naivety of the rich and the poor via a single parameter, $\frac{\beta}{\gamma} = a$, which describes the severity of the poor's naivety relative to the rich's. If $a = 1$ then both groups are "equally naive", if $a > 1$ then the poor are more naive than the rich. Using equation (11), I can investigate how $\hat{y}^*$ changes with $a$ (see computation in Appendix A.3):

$$\frac{d\hat{y}^*}{da} = \frac{(1 - F(\hat{y}^*))(\bar{E}(\hat{y}^*) - \hat{y}^*)}{a(1 - F(\hat{y}^*)) + F(\hat{y}^*)} > 0$$

The equilibrium cutoff $\hat{y}^*$ is increasing in the degree of naivety of the poor relative to the rich. The higher $a$, the more the poor tend to underestimate the benefits of sorting (relative to the rich) and hence the more they need to see of the whole distribution relative to the rich to have the same perceived benefits of sorting as the rich.

As naivety goes to zero, what happens to the equilibrium cutoff depends on the speed of convergence of $\beta$ resp. $\gamma$. If $\beta$ converges to zero faster than $\gamma$, $a$ goes to zero and $\hat{y}^*$ goes to 0. If $\gamma$ converges at a faster speed than $\beta$, $a$ converges to infinity and the equilibrium cutoff goes to $y_{\max}$.[16]

## 6.2 The consistency requirement

At this point it is instructive to look at the role of the consistency requirement in my model, using the specific example of this section. The equilibrium condition in this specifc example boils down to (9) and therefore the unique equilibrium cutoff $\hat{y}^* > 0$ needs to satisfy

$$\bar{E}(\hat{y}^*) - \underline{E}_r(\hat{y}^*) = \bar{E}_p(\hat{y}^*) - \underline{E}(\hat{y}^*),$$

i.e. the perceived difference in group average incomes needs to be the same for both the rich and the poor group in equilibrium. Figure 1 depicts the perceived group differences (in terms of average income) of the poor group (blue) and the rich group (red) as well as the correct benefits of sorting of the person at the cutoff (black) as a function of the cutoff $\hat{y}$ (for a truncated log-normal income distribution). For small $\hat{y}$, the rich perceive the difference between the two groups almost correctly, while the poor underestimate it a lot. This is because of the assumption I make on the bias: the larger the part of the income distribution that a group sees, the less biased they are about the other group.

[16] The best way to see the latter is to introduce the auxiliary parameter $b = \frac{\gamma}{\beta}$ in this case and rewrite $h(y)$ in terms of $b$.

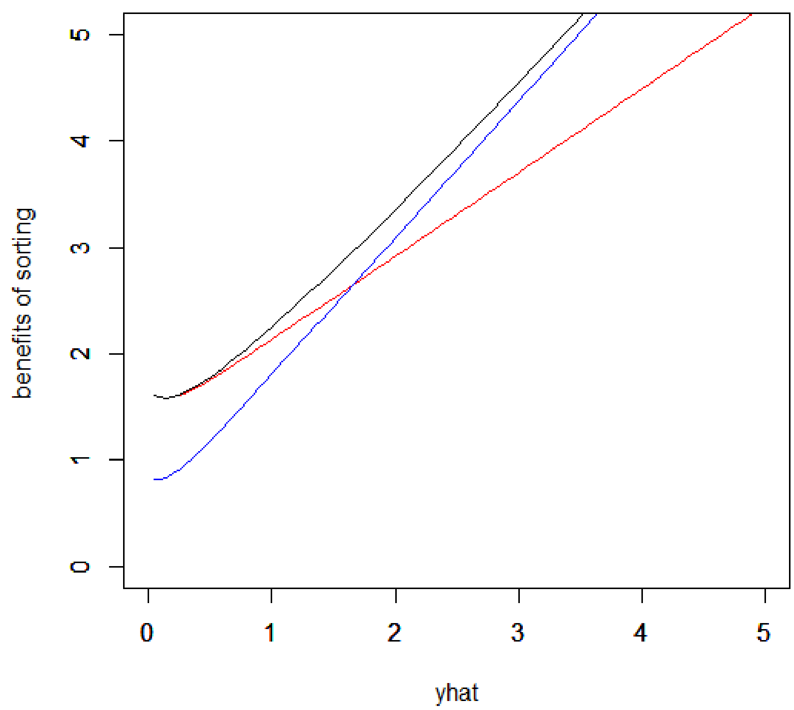


Figure 1: Perceived benefits of sorting for the rich (red) and poor (blue) and correct benefits of sorting as a function of the cutoff $\hat{y}$

This also implies that as $\hat{y}$ increases, the rich become more and more biased and the poor become more and more correct about the group difference. The blue and the red line cross at $\hat{y}^*$, the unique binary biased sorting equilibrium with consistency, where both groups have the same perceived benefits of sorting. As $\hat{y}$ increases beyond this point, the poor group starts to value sorting more than the rich group.

For a sorting equilibrium *without* consistency, the only condition that needs to be satisfied is that the cutoff is such that everybody in the rich group prefers being in the rich group to being in the poor group, while everybody in the poor group wants to stay in the poor group for some sorting fee $b > 0$. In Figure 1, all cutoffs $\hat{y}$ below $\hat{y}^*$ would satisfy this condition - if $\hat{y} \in (0, \hat{y}^*)$, the marginal person in the rich group values being in the rich group more than the marginal person in the poor group, and therefore we would be able to find a sorting fee $b > 0$ that the rich are willing to pay, while it doesn't seem worthwhile for the poor to do so. Hence, all $\hat{y} \in (0, \hat{y}^*]$ are binary biased sorting equilibria. Meanwhile, none of the $\hat{y}$ above $\hat{y}^*$ can be biased sorting equilibria, because the marginal person in the poor group would always be willing to pay more to join the rich group than the marginal person in the rich group, and thus no $b > 0$ could be found that separates the rich from the poor. Note however, that all $\hat{y} \in (0, \hat{y}^*)$, while being biased sorting equilibrium cutoffs, fail to satisfy the consistency requirement: The marginal person in the poor group, when looking at the people with just a little bit higher income than herself, cannot understand why these people want to be in the rich group and pay sorting fee $b = \hat{y}(\bar{E} - \underline{E}_r)$, because for her, being in the rich group is only worth $\hat{y}(\bar{E}_p - \underline{E})$, which is less.

In the specific case analyzed here, the consistency requirement selects a unique equilibrium out of the range of sorting equilibria. This is because the misperceptions converge to the truth monotonically, which implies that the blue line approaches the black line monotonically as the cutoff increases, while the red line approaches the black line monotonically as $\hat{y}$ decreases. Therefore, the two lines can only cut once. If the misperceptions were not monotone, the blue and the red line could be non-monotone, and therefore intersect several times. Each of those intersections would then constitute a biased sorting equilibrium with consistency. Consistency alone is not enough to guarantee uniqueness. Consistency and monotonicity of the misperceptions together do the job (for linear $U$).

Another way to interpret the consistency requirement is a refinement to "no-learning partitions". If a partition satisfies consistency, then people never come across anything that goes against their beliefs and surprises them, therefore they have no impulse to modify their beliefs or their actions in any way.

I do *not* model any form of learning in this paper. I also do not make any assumptions about what happens if people encounter other people, whose choices they do not understand. One possibility is that people just assume that the others are wrong, if they are puzzled by their choices, and do not modify their own beliefs or actions. Another possibility is, that they start to question their own beliefs about the other groups and maybe try to update them, based on choices of other people that they observe. Or they might even experiment

and join another group to learn about average income in that group. What the consistency requirement does is it restricts the set of biased sorting equilibria to those partitions, where neither of the above happens, because people are simply not puzzled by anybody else's choices. In that sense, the consistency requirement can be viewed as a stability refinement: consistent equilibrium partitions are stable with respect to learning, experimenting or updating. Because what they see is consistent with their beliefs about the world, people have no incentives to question or change their beliefs, irrespective of what they would do if they would encounter anything that is at odds with their beliefs.

## 6.3 An Application to Segregation and Preferences for Redistribution

Economic segregation can exacerbate inequalities in various ways. Schooling is one prominent example: If children living in affluent areas get better education than children from poor neighbourhoods because their local schools are of a better standard due to high local investment, income inequality in the next generation will be amplified. This effect is specifically pronounced in the United States, where school choice is linked to neighbourhood (see e.g. Chetty et al. (2014)). Moreover, having class mates coming from rich and influential families might not only have the direct effect on education via better quality of schooling, but might also yield benefits later in life through social connections that lead to better jobs and opportunities (see e.g. Savage (2015)).

In this section, I demonstrate that there might be another channel through which segregation can affect economic inequality: Economic segregation, if it leads to misperceptions of the income distribution, can have significant consequences for support for redistribution in society, and hence for (post-tax and post-redistribution) income inequality. I show that segregation leads poor people to underestimate what they can gain from redistribution and therefore to show less support for redistribution than if they would have perfect knowledge of the income distribution. Moreover, an increase in inequality (in the form of a mean-preserving spread of the income distribution) always leads to a smaller increase in perceived inequality and therefore in the demand for redistribution than if people were unbiased. The reason for this is that people with income below average fully observe the fall of low incomes, but do not fully see the offsetting increase of high incomes. Therefore, they think that average income has decreased. But because people's gains from redistribution depend positively on the difference between their own income and (perceived) average income, and both decrease if people are biased, demand for redistribution increases less than if people are unbiased and know that average income hasn't changed. I show that the increase in inequality can even be such that perceived inequality declines and therefore people's support for redistribution falls.

In the following analysis, I continue to use the functional forms of $\bar{E}_p(\hat{y})$ and $\underline{E}_r(\hat{y})$ as specified in (7) and (8), because this enables me to derive precise results. However, the general flavour of those results would not change if more general specifications of $\bar{E}_p(\hat{y})$ and $\underline{E}_r(\hat{y})$ were used that satisfy the conditions

for existence of a unique equilibrium above the median, given in Appendix A.2.

### 6.3.1 Inequality and the demand for redistribution

Suppose that everybody in the economy has to pay a proportional tax $t$ and the government redistributes the proceeds equally among all its citizens afterwards. Hence, a person with pre-tax income of $y_i$ has after-tax and after-redistribution income

$$(1-t)y_i + \tau(t)E$$

where the function $\tau(t) \leq t$ accounts for the fact that there is a deadweight loss of taxation. (And let $\tau(.)$ be such that $\tau(t) > 0 \; \forall t \in (0,1)$, $\tau(0) = 0$, $\tau''(t) \leq 0$, $\tau(1) = 0$, $\tau'''(t) \geq 0$ [this guarantees that $\tau'(t)$ is convex and hence also $\tau'^{-1}$ is convex, given that $\tau'$ is decreasing]). Suppose furthermore that people vote to decide on the tax rate, and suppose that they care only about their own post-tax income.

Meltzer and Richard (1981) have examined the relationship between inequality and the demand for redistribution in this model: If people are unbiased about the income distribution, when voting for the redistribution rate a person with income $y_i$ will simply choose the tax rate $t$ that maximizes her post-tax income

$$(1-t)y_i + \tau(t)E.$$

As preferences are single-peaked in this case, the tax rate determined by majority voting will be the median earner's optimal tax rate given by

$$\tau'(t^*) = \frac{y^M}{E}$$

if $\frac{y^M}{E} \leq 1$ and $t^* = 0$ otherwise. As $\tau'(t)$ is decreasing in $t$, the decisive voter's optimal tax rate $t^*$ is decreasing in the ratio between median and average income.

The ratio $\frac{y^M}{E}$ can be regarded as an, albeit rudimentary, measure of the degree of income equality in society. If the ratio is small, this means the difference between median and mean income is large and the income distribution has a large positive skew with a majority of people earning income below average and a few very rich people. Therefore, income equality is low and the demand for redistribution will be high in that case. If, on the other hand, the income distribution is almost symmetric, with most people being middle-class and only a few at the bottom and the top of the distribution, the equality ratio $\frac{y^M}{E}$ will be large (i.e. close to 1), and demand for redistribution will be low.

To analyze people's preferences for redistribution if they are biased, I need to establish what their perception of average income is: If people would correctly perceive both average income in their group and average income in the other group, they could simply calculate overall average income via the formula

$$E = F\underline{E}(\hat{y}) + (1-F)\bar{E}(\hat{y})$$

for any cutoff $\hat{y}$.[17] However, if there is economic segregation and people are biased, then people misperceive average income in the other group, and hence they mis-estimate overall average income. Specifically, poor people think that average income is

$$E_p(\hat{y}) = F\underline{E}(\hat{y}) + (1-F)\bar{E}_p(\hat{y}) < E.$$

Because they underestimate average income in the rich group,

$$\bar{E}_p(\hat{y}) < \bar{E},$$

they end up underestimating overall average income. Analogously, rich people overestimate average income,

$$E_r = F\underline{E}_r(\hat{y}) + (1-F)\bar{E}(\hat{y}) > E.$$

Let me for simplicity of exposition assume henceforth that rich and poor people are equally naive, i.e. $\beta = \gamma$,[18] and remember that in this case the equilibrium cutoff will always be at average income $E$. This implies that the median earner is in the poor group (because the income distribution is positively skewed) and her preferred tax rate is given by

$$\tau'(\tilde{t}^*) = \frac{y^M}{E_p(E)} \text{ (or } \tilde{t}^* = 0 \text{ if } E_p(E) < y^M)$$

$E_p$ is smaller than $E$, hence the median earner's perceived degree of equality as measured by $\frac{y^M}{E_p(E)}$ is higher than without segregation. Therefore, her optimal tax rate is lower in the presence of economic segregation.

**Lemma 1** *In the model with segregation and misperceptions the median earner's preferred tax rate is lower compared to the model without misperceptions.*

For the remainder of this paper, I will assume that the following condition on the income distribution and people's naivity holds:

**Assumption 2** *The distance between median and mean income is sufficiently high, such that*

$$\frac{E}{E_r(E)} \geq \frac{y^M}{E_p(E)}.$$

[17] Note that I assume that people know the relative size of their respective group, i.e. they know $F(\hat{y})$ and $1 - F(\hat{y})$. They also know the range of the distribution and where the cutoff lies. They only misperceive the shape of the distribution function in the other group. With the type of bias that I examine here, their perceived income distribution in the other group is more skewed towards $\hat{y}$ compared to the actual distribution.

[18] The analysis can be done in a similar way for the general case of $\beta \neq \gamma$.

**Remark 1** *In Appendix A.5 I show that* $\frac{E}{E_r(E)} \geq \frac{y^M}{E_p(E)}$ *is guaranteed for misperceptions (7) and (8) if*

$$E - y^M \geq \beta \frac{\bar{E}(E) - \underline{E}(E)}{4}$$

*This condition holds if* $E - y^M$ *is large enough compared to* $\bar{E}(E) - \underline{E}(E)$*, i.e. if the distribution is positively skewed but there is not too much mass at the tails of the distribution, and if* $\beta$ *is small, i.e. people are not too biased.*[19]

**Lemma 2** *If Assumption A.5 holds, the median earner is the decisive voter.*

The preferred tax rate of the poorest person in the rich group (i.e. the person earning average income $E$) is given by

$$\tau'(t) = \frac{E}{E_r(E)}$$

If the distance between median and mean income is sufficiently high, such that Assumption 2 holds, then this person will demand a lower tax rate than the median earner, and hence the median earner will be the decisive voter. As the median earner wants less redistribution than in the unbiased case, the tax rate selected by majority voting will be lower and therefore demand for redistribution in this segregated society will be lower than in a society without segregation and misperceptions.

**Proposition 8** *The tax rate selected by majority voting in a segregated society where people misperceive the shape of the income distribution as described above is lower than in a society without segregation and misperception of the income distribution.*

**Proof.** See above. ■

## 6.4 The effect of changing inequality on demand for redistribution

In the following section I analyze what happens to people's (mis)perceptions and the support for redistribution in a segregated society if income inequality increases and how the effects differ compared to a society without segregation. When analyzing the effect of an increase in inequality, it is important to clearly specify the exact form of this increase in inequality. Some changes in the shape of the income distribution are such that it cannot even be unequivocally decided whether they lead to an increase or decrease in inequality - different measures of

[19] Assumption 2 holds for positively skewed income distributions that look like actual income distributions that we observe in the real world, for example it holds for a truncated lognormal (on $(0, 10^8)$ ) with $\sigma = 1.2$ and $\mu = 11$ (the US household income distribution can be approximated by this function), and equally for a scaled down version of it, a truncated lognormal on $(0, 10)$ with $\mu = 0.8$ and $\sigma = 1$ (both times $\beta = 0.1$).

inequality might yield different results. However, any mean-preserving spread of the income distribution always implies an increase in inequality, irrespective of the measure that is used, because it can be decomposed into (potentially infinitely many) transfers between rich and poor where money is transferred from a relatively poor to a relatively rich person. It therefore increases all measures of inequality that respect the *principle of transfers*, such as the Gini coefficient or the Theil index (see also Cowell (2000) and Dalton (1920)).[20] Hence, I will focus on the effect of a mean-preserving spread of the income distribution on group formation and demand for redistribution.

For simplicity, I require the mean-preserving spread to be such that the mass of people below and above the mean remain the same, but mass shifts from the middle towards the endpoints of the distribution, such that median income declines.[21] Specifically, I will analyze the effect of a mean-preserving spread of the income distribution such that $\bar{E}$ increases, $\underline{E}$ goes down and $F(E)$ remains unchanged. Furthermore, I require Assumption 2 to hold before and after the change in inequality. As this implies that the median earner is always the median voter, I will use these two expressions interchangeably.

In the absence of segregation and misperceptions, the median voter's support for redistribution increases due to a mean-preserving spread of the above described form, because median income declines relative to average income and hence the equality ratio $\frac{y^M}{E}$ decreases,

$$\Delta\left(\frac{y^M}{E}\right) = \frac{\Delta y^M}{E} = \frac{\Delta y^M}{y^M}\frac{y^M}{E},$$

i.e. the percentage change in $\frac{y^M}{E}$ is $\frac{\Delta y^M}{y^M}$ (where $\Delta y^M < 0$). This means that demand for redistribution, given by

$$\tau'(t^*) = \left(\frac{y^M}{E}\right),$$

increases. The increase in the median voter's optimal tax rate $t^*$ is

$$\Delta t^* = \tau'^{-1}\left(\frac{y^M + \Delta y^M}{E}\right) - \tau'^{-1}\left(\frac{y^M}{E}\right)$$

In a segregated society, where people misperceive the shape of the income distribution, the effect of an increase in inequality on the support for redistribution depends on its impact on the location of the equilibrium cutoff $\hat{y}^*$, because this determines people's beliefs about the other group's average income. Recall that the equilibrium cutoff $\hat{y}^*$ is the fixed point of the function

$$h(\hat{y}) = \frac{a(1 - F(\hat{y}^*))\bar{E}(\hat{y}^*) + F(\hat{y}^*)\underline{E}(\hat{y}^*)}{a(1 - F(\hat{y}^*)) + F(\hat{y}^*)}$$

[20] In the economic literature, any inequality measure is generally required to satisfy four properties: anonymity, scale independence, population independence and the principle of transfers. For an extensive discussion of different inequality measures see Cowell (2000).

[21] This implies that the distance between mean and median income increases.

As described in Section 6, $h(\hat{y})$ has a unique fixed point, which is at average income $E$ if $a = 1$. Hence, the position of the equilibrium cutoff does not change due to a mean-preserving spread.

What happens to perceived inequality and the demand for redistribution? As I explained in the previous section, if people are biased due to segregation, the median voter's optimal tax rate $\tilde{t}^*$ is characterized by the equation

$$\tau'(\tilde{t}^*) = \left(\frac{y^M}{E_p(E)}\right)$$

where $\tilde{t}^* < t^*$ (because $E_p < E$) - the median earner's preferred tax rate is lower under segregation because perceived equality $\frac{y^M}{E_p}$ is higher. While average income $E$ does not change due to a mean-preserving spread, I show in Appendix A.6 that average perceived income of the poor, $E_p$, declines. The poor feel that average income declines because they experience the decline of average income in their own group fully, but only partially take note of the compensating increase in average income among the rich. Hence, they think that society as a whole has become poorer. As a result, the change in the perceived equality ratio $\frac{y^M}{E_p}$ amounts to

$$\Delta\left(\frac{y^M}{E_p}\right) = \frac{\Delta y^M E_p - y^M \Delta E_p}{(E_p)^2} = \left(\frac{\Delta y^M}{y^M} - \frac{\Delta E_p}{E_p}\right)\frac{y^M}{E_p}$$

and thus the percentage decrease in $\frac{y^M}{E_p}$ is $\frac{\Delta y^M}{y^M} - \frac{\Delta E_p}{E_p}$, which is smaller (in absolute terms) than the percentage decrease of $\frac{y^M}{E}$ in the unbiased case, because $\frac{\Delta E_p}{E_p} < 0$.

**Proposition 9** *If society is segregated, an increase in inequality (in the form of a mean-preserving spread) always leads to a smaller percentage decrease in the median voter's perceived equality than in the absence of segregation and misperception.*

Moreover, in Appendix A.6 I demonstrate that one can always construct a mean-preserving spread that leads the median voter to believe that society has become more rather than less equal, i.e that inequality has decreased rather than increased.

**Proposition 10** *There exists an increase in inequality that causes a decrease of the median earner's perceived degree of inequality under segregation.*

The intuition for Proposition 10 is that, unlike in the non-segregated case, the median voter's perceived equality ratio $\frac{y^M}{E_p}$ can increase due to a mean preserving spread if people are biased, because both $y^M$ and $E_p$ decline. If the mean-preserving spread is such that the median voter's perceived degree of inequality decreases, as in Proposition 10, then also the median voter's demand for redistribution (i.e. her preferred tax rate) must necessarily decrease.

**Corollary 6** *There always exists an increase in inequality such that the tax rate determined by majority voting decreases under segregation.*

In Appendix A.6, I derive the condition on the mean-preserving spread that guarantees Proposition 10. As I explain above, this condition must ensure that the decline in $E_p$ is larger than the decline in $y^M$. I also derive a weaker condition on the mean-preserving spread that guarantees that even though perceived inequality does not decrease, demand for redistribution increases less under segregation than without segregation. The step-by-step calculations in Appendix A.6 can be summarized as follows: If perceived equality decreases due to a mean-preserving spread under segregation, the fact that the percentage decrease in perceived equality is smaller if society is segregated is not enough to guarantee that also the increase in demand for redistribution will be smaller than without segregation. There are two reasons for this: First, as perceived equality is higher to start with under segregation, a smaller percentage decrease does not automatically imply a smaller absolute decrease than in the absence of segregation. Second, even if the decrease in perceived equality is lower also in absolute terms, it is not clear whether the increase in demand for redistribution will be lower as well: this depends on the shape of the deadweight loss function $\tau(.)$. However, it turns out that the assumption that $\tau'$ is decreasing and convex is sufficient to ensure that demand for redistribution increases less under segregation if the absolute decrease in perceived equality is smaller than in the absence of segregation. The condition on the mean-preserving spread that guarantees that demand for redistribution under segregation increases by less if inequality increases compared to a situation without segregation is weaker than the condition that is needed for Proposition 10.

In Appendix A.8, I describe how more general changes in the shape of the income distribution affect demand for redistribution if society is segregated.

Another way to model the decline in perceived inequality after an increase in inequality is to assume that there is no segregation in place before the change (because whoever offers the sorting technology doesn't find it worthwhile) but then as inequality increases, offering the sorting technology becomes profitable and therefore society becomes segregated (and people become biased). I examine this in the following subsection for the case of a profit-maximizing monopolist.

## 6.5 Inequality and the supply side of sorting

Who is offering the possibility to sort into groups? Suppose a profit-maximizing monopolist, who has a fixed cost $c > 0$ of offering the sorting technology, can decide whether or not to become active.[22] Her profits from offering sorting are

$$\hat{y}^*(\bar{E}(\hat{y}^*) - \underline{E}_r(\hat{y}^*))(1 - F(\hat{y}^*)) - c$$

[22] In Appendix A.11, I show that the argument works in the same way if it is a welfare-maximizing social planner who decides about offering the sorting technology.

Given that the equilibrium cutoff is at $E$ and substituting for $\underline{E}_r$, this can be rewritten as

$$E(E - \underline{E}(E))[1 - \gamma F(E)(1 - F(E))] - c \tag{12}$$

Suppose that initially the income distribution is such that

$$E(E - \underline{E}(E))[1 - \gamma F(E)(1 - F(E))] - c < 0$$

and hence the monopolist would prefer to stay out of the market. If inequality increases (again in the sense of a mean-preserving spread of the income distribution, which decreases $\underline{E}$ and increases $\bar{E}$, while leaving $E$ and $F(E)$ constant), $E - \underline{E}$ increases. This means that if the increase in inequality is sufficiently large, the profits from offering the sorting technology will become positive and the society will become segregated. Thus, a large enough increase in inequality will lead to economic segregation.[23]

**Lemma 3** *Suppose that the income distribution is initially such that a profit maximizing monopolist with fixed costs $c > 0$ does not find it profitable to offer the sorting technology. Then for any $c > 0$ there exists a mean-preserving spread of the income distribution such that the monopolist's profits become positive.*

Hence, I can compare the effect of increasing inequality in the presence of segregation to its effect without taking into account segregation (and the resulting misperception). As in the previous sections, I require Assumption 2 to be satisfied after the increase in inequality, to ensure that the median earner is the decisive voter.

If inequality increases then if there is no segregation and people are unbiased, the median voter will demand more redistribution than before the change, because median income $y^M$ is smaller as a result of the mean-preserving spread, and hence also $\frac{y^M}{E}$ decreases:

$$\Delta\left(\frac{y^M}{E}\right) = \frac{\Delta y^M}{E}$$

Therefore, the median earner's demand for redistribution increases from

$$\tau'^{-1}\left(\frac{y^M}{E}\right)$$

to

$$\tau'^{-1}\left(\frac{\dot{y}^M}{E}\right)$$

where $\dot{y}^M = y^M + \Delta y^M < y^M$ is median income after the increase in inequality.

[23] The increase in inequality need not be necessarily in the sense of a mean-preserving spread for the result to hold. In Appendix A.4, I show that in the case of a log-normal income distribution, an increase in the log-variance $\sigma$ (which is equivalent to an increase in the Gini-coefficient) also leads to an increase in the monopolist's revenue.

If the increase in inequality leads to economic segregation and hence causes people to be biased, then the median voter's demand for redistribution changes from

$$\tau'^{-1}\left(\frac{y^M}{E}\right)$$

to

$$\tau'^{-1}\left(\frac{\dot{y}^M}{E_p(E)}\right)$$

where

$$E_p(E) = E - \beta(1 - F(E))^2(\bar{E}(E) + \Delta\bar{E}(E) - E)$$

As $E_p < E$, the increase in the median voter's demand for redistribution will be smaller than in the absence of economic segregation.

**Proposition 11** *If an increase in inequality leads to economic segregation, the median voter's demand for redistribution will increase less than in the absence of segregation.*

In Appendix A.7 I show that I can always construct a mean preserving spread of the income distribution such that demand for redistribution decreases under segregation.

**Proposition 12** *There exists an increase in inequality that causes economic segregation and leads to a decline in the tax rate determined by majority voting.*

**Remark 2** *Apart from the mean-preserving spread described above there are also other types of increases in inequality that would make it profitable for the monopolist to offer one cutoff. I demonstrate in Appendix A.4 that for the lognormal distribution an increase in $\sigma$ (which corresponds to an increase in the Gini-coefficient but is a median-preserving instead of a mean-preserving spread) also leads to an increase in the monopolist's profits (12).*

## 6.6 Empirical Evidence

In February 2016, I conducted an online survey on 600 US citizens above the age of 18. The survey was distributed via Amazon Mechanical Turk and the original questionnaire can be accessed at `https://lse.ut1.qualtrics.com/jfe/form/SV_eDLNkeGfQg2ycM5`. A description of the sample (i.e. respondents' characteristics) can be found in Appendix A.9.[24] The advantages and potential pitfalls of using Amazon Mechanical Turk in academic research have been discussed by Kuziemko et al. (2015) in their Online Appendix. I summarize some of their points and document my own experiences in Appendix A.9.1.

[24] The data and all do-files are available upon request.

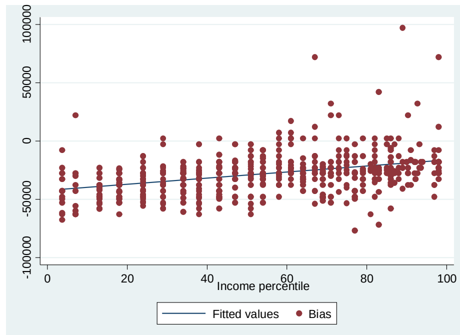


Figure 2: People's estimate of average income is increasing in their own income (Bias = correct average income - perceived average income).

By conducting this survey, I wanted to address two main questions:

1. Is there evidence that people misperceive the income distribution in the way I assume in the above application of the model to the question of inequality and redistributive preferences? For example, do poor people underestimate overall average income and do rich people overestimate it?

2. Are people with a diverse social circle (i.e. people who are not very "segregated") less biased?

To tackle the first question, I asked people about their own household income and their estimate of average US household income. Figure 2 plots the relationship between the two: in general, both rich and poor people underestimate mean household income on average. However, people's estimate of average household income is increasing in their own income (see also Figure 10 in the Appendix). This is roughly in line with my model, which would predict that poor people underestimate average income (because they know average income in their group and underestimate average income in the rich group) and rich people overestimate average income (because they know their own average income and overestimate the poor group's income).

The first attempt to identify a link between segregation and misperception is to look at the relationship between the degree of income segregation that a respondent lives in and (the absolute value of) her bias. For this purpose, I match the survey data with county-level income segregation data computed by Chetty et al. (2014). However, I do not find any relationship between county-level income segregation and a respondent's absolute level of bias. I suspect that county-level data is too coarse to be useful as a proxy for an individual's degree of segregation. Unfortunately, I cannot repeat the analysis with a more precise measure of income segregation because I have neither lower-level locational information about my respondents, nor data on lower-level income segregation in the US.

Table 1: Misperceptions of average household income are less severe if respondents have more diverse social circles.

| | Bias |
|---|---|
| Income percentile | $0.004^{***}$ |
| | (0.001) |
| (Income percentile) x (Social segregation) | $0.002^{**}$ |
| | (0.001) |
| Social segregation | -0.073 |
| | (0.060) |
| Intercept | $-0.598^{***}$ |
| | (0.041) |
| $N$ | 592 |

$p$-values in parentheses

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

However, I also tried to elicit respondents' individual degrees of segregation by asking about the diversity of their social interactions. In particular, I asked about their friends and colleagues, and how many of them have similar resp. different levels of household income and education. Then I employed a scale from 0 to 4 to classify respondents as more or less segregated (4 indicating the highest possible degree of segregation) concerning those social circles, depending on how similar their work colleagues resp. friends are to themselves. Subsequently, I used factor analysis to identify a common factor out of these categorical response variables (for detailed explanations see Appendix A.9).

I find that the misperception of average income is amplified by social segregation: poor people underestimate average household income less and rich people overestimate it less if their social circle is more diverse (see Table 1).

Furthermore, I asked the so-called "Lin position generator" question in the version of the "Great British Class calculator"[25], which is the short version of a similar question asked in the Great British Class Survey (see Savage (2015)).[26] This question tries to identify the diversity of the respontent's social circle by asking whether she socially knows people with certain occupations (eighteen different occupations), ranging from chief executive to cleaner. I measure diversity of the social circle by assigning to each of the occupations their status rank using the Cambridge Social Interaction and Stratification (CAMSIS) scale score (where low numbers correspond to high rank) and then calculating for each respondent the standard deviation of all the scores of occupations she knows: the higher this standard deviation, the more diverse can the respondent's social circle be asssumed to be. Regressing the absolute value of people's misperception

[25] see `http://www.bbc.co.uk/news/magazine-22000973`

[26] The question is named after the sociologist Nan Lin who developed it in the 1980s.

of average income in percentage terms (variable *Bias2*) on the standard deviation yields significant results and the coefficient has the expected sign: A more diverse social circle corresponds to less bias about average household income (see Table 2).

Table 2: Misperceptions of average income are amplified by social segregation.

| | (1) | (2) |
|---|---|---|
| | Bias2 | Bias2 |
| Social circle status diversity | -0.0107$^{***}$ | -0.00916$^{**}$ |
| | (0.005) | (0.015) |
| Income percentile | | -0.00181$^{***}$ |
| | | (0.000) |
| Intercept | 0.483$^{***}$ | 0.568$^{***}$ |
| | (0.000) | (0.000) |
| $N$ | 592 | 592 |

$p$-values in parentheses

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

# 7 Conclusion

In this paper I have presented a model of sorting in the presence of misperceptions that takes into account the two-way interaction between beliefs about society and social segregation. I have defined a new equilibrium concept, the biased sorting equilibrium, which can be interpreted as characterizing partitions which are stable given the misperceptions that people acquire once they are segregated and interact only with people in their own social group. I have also introduced a refinement concept, the consistency requirement, and I have showed that it guarantees that the biased sorting equilibrium partitions will be monotone. Moreover, it guarantees uniqueness of the equilibrium partition if people's misperceptions are such that they converge to the truth monotonically as the size of their group increases.

I have showed how this general framework can be applied to the issue of income inequality and preferences for redistribution: If people are segregated according to income, there will be less demand for redistribution in society. Furthermore, an increase in inequality will lead to a smaller increase in support for redistribution than in the absence of segregation, and certain mean-preserving spreads of the income distribution can even lead to a decrease in demand for redistribution, because they result in a decline in perceived inequality.

Finally, I have reported some of my empirical findings on misperception of the shape of the income distribution and segregation: I have showed evidence that people's estimate of average household income is increasing in their own

income, and that people's misperceptions are more severe, the more socially segregated they are.

My approach shows that modelling segregation and belief formation simultaneously can yield interesting and unexpected results and offers new perspectives on issues such as income inequality and redistribution. In the present paper, I have used the model to examine the implications of segregation and biased beliefs on redistributive demand, but the general framework that I have developed in this paper offers itself to a wide set of applications related to segregation, such as education policy and housing.

# References


[1] Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The elite illusion: Achievement effects at Boston and New York exam schools. Econometrica, 82(1), 137-196.

[2] Alesina, A., & Angeletos, G. M. (2005). Fairness and Redistribution. American Economic Review, 960-980.

[3] Algan, Y., Do, Q. A., Dalvit, N., Le Chapelain, A., & Zenou, Y. (2015). How Social Networks Shape Our Beliefs: A Natural Experiment among Future French Politicians. (mimeo)

[4] Andreasch, M., Fessler, P., Mooslechner, P., Schürz, M. (2013) Fakten zur Vermögensverteilung in Österreich, BMASK (Ed.), Sozialbericht 2011-2012 (Vienna 2013) 247-265.

[5] Ashok, V., Kuziemko, I., & Washington, E. (2015). Support for Redistribution in an Age of Rising Inequality:: New stylized facts and some tentative explanations. Brookings papers on economic activity, (1), 2.

[6] Bartels, L. M. (2009). Economic inequality and political representation. The unsustainable American state, 167-96.

[7] Becker, G. S. (1974). A theory of marriage. In Economics of the family: Marriage, children, and human capital (pp. 299-351). University of Chicago Press.

[8] Benabou, R., & Ok, E. A. (2001). Social Mobility and the Demand for Redistribution: The Poum Hypothesis. The Quarterly Journal of Economics, 116(2), 447-487.

[9] Boisjoly, J., Duncan, G. J., Kremer, M., Levy, D. M., & Eccles, J. (2006). Empathy or antipathy? The impact of diversity. The American economic review, 96(5), 1890-1905.

[10] Bonica, A., McCarty, N., Poole, K. T., & Rosenthal, H. (2013). Why hasn't democracy slowed rising inequality?. The Journal of Economic Perspectives, 27(3), 103-123.

[11] Borge, L., & Rattsoe, E. J. (2004). Income Distribution and Tax Structure: Empirical Test of the Meltzer Richard Hypothesis. European Economic Review, 48, 805-826.

[12] Burns, J., Corno, L., & La Ferrara, E. (2013). Does interaction affect racial prejudice and cooperation? Evidence from randomly assigned peers in South Africa. (mimeo)

[13] Chetty, R., Hendren, N., Kline, P., & Saez, E. (2014). Where is the land of opportunity? The geography of intergenerational mobility in the United States (No. w19843).

[14] Cowell, F. A. (2000). Measurement of inequality. Handbook of income distribution, 1, 87-166.

[15] Cowell, F. (2011). Measuring inequality. OUP Oxford.

[16] Cruces, G., Perez-Truglia, R., & Tetaz, M. (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. Journal of Public Economics, 98, 100-112.

[17] Dalton, H. (1920). The measurement of the inequality of incomes. The Economic Journal, 30(119), 348-361.

[18] Damiano, E., & Li, H. (2007). Price discrimination and efficient matching. Economic Theory, 30(2), 243-263.

[19] Dustmann, C., & Preston, I. (2001). Attitudes to ethnic minorities, ethnic context and location decisions. The Economic Journal, 111(470), 353-373.

[20] Evans, G. & Tilley, J. (2011). Private schools and public divisions: The influence of private schooling on social attitudes and political choices. in Alison Park et al (eds.), British Social Attitudes: The 28th Report. London: Sage.

[21] Georgiadis, A., & Manning, A. (2012). Spend it like Beckham? Inequality and redistribution in the UK, 1983–2004. Public choice, 151(3-4), 537-563.

[22] Golub, B., & Jackson, M. O. (2012). How Homophily Affects the Speed of Learning and Best-Response Dynamics. The Quarterly Journal of Economics, 127(3), 1287-1338.

[23] Gouveia, M., & Masia, N. A. (1998). Does the median voter model explain the size of government?: Evidence from the states. Public Choice, 97(1-2), 159-177.

[24] Karadja, M., Mollerstrom, J., & Seim, D. (2014). Richer (and Holier) than Thou? The Effect of Relative Income Improvements on Demand for Redistribution (No. 1050). George Mason University, Interdisciplinary Center for Economic Science.

[25] Kenworthy, L., & McCall, L. (2008). Inequality, public opinion and redistribution. Socio-Economic Review, 6(1), 35-68.

[26] Kiatpongsan, S., & Norton, M. I. (2014). How much (more) should CEOs make? A universal desire for more equal pay. Perspectives on Psychological Science, 9(6), 587-593.

[27] Kuziemko, I., Norton, M. I., Saez, E., & Stantcheva, S. (2015). How elastic are preferences for redistribution? Evidence from randomized survey experiments. The American Economic Review, 105(4), 1478-1508.

[28] Larcinese, V. (2005). Electoral competition and redistribution with rationally informed voters. Contributions in Economic Analysis & Policy, 4(1).

[29] Levy, G. and Razin, R. (2016). The Coevolution of Segregation, Polarised Beliefs and Discrimination: The Case of Private vs. State Education (mimeo)

[30] Levy, G., & Razin, R. (2015). Preferences over equality in the presence of costly income sorting. American Economic Journal: Microeconomics, 7(2), 308-337.

[31] Matakos, K. & Xefteris, D. (2016) Divide and Rule: Redistribution in a Model with Differentiated Candidates. forthcoming in Economic Theory

[32] Meltzer, A. H., & Richard, S. F. (1981). A rational theory of the size of government. The Journal of Political Economy, 914-927.

[33] Meltzer, A. H., & Richard, S. F. (1983). Tests of a rational theory of the size of government. Public Choice, 41(3), 403-418.

[34] Milanovic, B. (2000). The median-voter hypothesis, income inequality, and income redistribution: an empirical test with the required data. European Journal of Political Economy, 16(3), 367-410.

[35] Norton, M. I., & Ariely, D. (2011). Building a better America—One wealth quintile at a time. Perspectives on Psychological Science, 6(1), 9-12.

[36] Pesendorfer, W. (1995). Design innovation and fashion cycles. The American Economic Review, 771-792.

[37] Piketty, T. (1995). Social mobility and redistributive politics. The Quarterly journal of economics, 551-584.

[38] Rayo, L. (2013). Monopolistic Signal Provision. The BE Journal of Theoretical Economics, 13(1), 27-58.

[39] Rodriguez, F. (1999). Inequality, redistribution and rent-seeking. Department of Economics, University of Maryland, 1999b.(mimeo).

[40] Saez, E., & Zucman, G. (2016). Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data. The Quarterly Journal of Economics, 131(2), 519-578.

[41] Savage, M. (2015). Social class in the 21st century. Penguin UK.

[42] Scervini, F. (2012). Empirics of the median voter: democracy, redistribution and the role of the middle class. The Journal of Economic Inequality, 10(4), 529-550.

[43] Wenglinsky, H. (2007). Are private high schools better academically than public high schools?[electronic resource]. Center on Education Policy, 2007.

[44] Windsteiger, L. (2016). Optimal signal provision and income inequality (mimeo)

# A Appendices

## A.1 Consistency and monotonicity

Without imposing the consistency requirement, also non-monotone partitions can be biased sorting equilibria (if the belief function is of a certain form): Suppose that $y_1 \in S_b$ and $y_2 \in S_0$ with $y_1 < y_2$. In order for the partition $[S_0, S_b]$ to constitute a biased sorting equilibrium, it must be the case that

$$U(y_1, E_b[S_0]) \leq U(y_1, E[S_b]) - b$$

and

$$U(y_2, E[S_0]) \geq U(y_2, E_0[S_b]) - b$$

(Notation: $E_i[S_j]$ is group $S_i$'s belief about average income in $S_j$.) Combined, these two conditions give

$$U(y_2, E_0[S_b]) - U(y_2, E[S_0]) \leq b \leq U(y_1, E[S_b]) - U(y_1, E_b[S_0])$$

It is immediate to see that whether this inequality can hold depends on the belief function, because even though $y_1 < y_2$, the misperceptions $E_0[S_b]$ and $E_b[S_0]$ could be defined in such a way that

$$U(y_1, E[S_b]) - U(y_1, E_b[S_0]) \geq U(y_2, E_0[S_b]) - U(y_2, E[S_0])$$

However, the consistency requirement rules out non-monotone equilibrium partitions for any belief function.

**Proposition 13** *All biased sorting equilibria with consistency satisfy monotonicity.*

**Proof.** Suppose a non-monotone equilibrium exists. Then it must be the case that there exist $y_1 \in S_b$ and $y_2 \in S_0$ with $y_1 < y_2$. Then the IC constraint for $y_1$ requires that

$$U(y_1, E_b[S_0]) \leq U(y_1, E[S_b]) - b$$

The consistency requirement additionally requires that

$$U(y_2, E_b[S_0]) \geq U(y_2, E[S_b]) - b$$

But these two conditions combined give

$$U(y_1, E[S_b]) - U(y_1, E_b[S_0]) \geq U(y_2, E[S_b]) - U(y_2, E_b[S_0])$$

which cannot hold for any belief function $B$ if $y_1 < y_2$ and $U$ is strictly supermodular. ■

## A.2 Necessary and sufficient conditions for existence and uniqueness of a biased sorting equilibrium with consistency

In this section I will derive necessary and sufficient conditions for existence and uniqueness of a biased sorting equilibrium with consistency in the four cases mentioned in Section 5.

The conditions for a biased sorting equilibrium can be written as

$$\begin{aligned} U(y, \bar{E}_p(\hat{y}^*)) - U(y, \underline{E}(\hat{y}^*)) &< b \ \ \forall y < \hat{y}^* && (13) \\ U(y, \bar{E}(\hat{y}^*)) - U(y, \underline{E}_r(\hat{y}^*)) &\geq b \ \ \forall y \geq \hat{y}^* && (14) \end{aligned}$$

for some $b > 0$. Due to supermodularity of $U$, these two conditions can be simplified to

$$\begin{aligned} U(\hat{y}^*, \bar{E}_p(\hat{y}^*)) - U(\hat{y}^*, \underline{E}(\hat{y}^*)) &\leq b \\ U(\hat{y}^*, \bar{E}(\hat{y}^*)) - U(\hat{y}^*, \underline{E}_r(\hat{y}^*)) &\geq b \end{aligned}$$

which implies that at the equilibrium cutoff we must have

$$U(\hat{y}^*, \bar{E}(\hat{y}^*)) - U(\hat{y}^*, \underline{E}_r(\hat{y}^*)) \geq U(\hat{y}^*, \bar{E}_p(\hat{y}^*)) - U(\hat{y}^*, \underline{E}(\hat{y}^*)) \qquad (15)$$

(and note that as $U(\hat{y}^*, \bar{E}(\hat{y}^*)) - U(\hat{y}^*, \underline{E}_r(\hat{y}^*)) \geq b$ the RHS of this inequality must be strictly positive in any biased sorting equilibrium).

Inequality (15) says that at any equilibrium cutoff it must be the case that the rich group's perceived benefit of sorting (LHS) is greater than the poor group's perceived benefit of sorting (RHS). It follows immediately that we cannot find a $b > 0$ such that a biased sorting equilibrium exists at any cutoff in Case 4, in which the rich group underestimates the difference between groups for any cutoff and the poor group overestimates the difference between groups for any cutoff. The reason is that the fact that $U$ is strictly increasing in both arguments implies that the rich's perceived benefit of sorting lie below the poor's benefit of sorting for every cutoff $\hat{y}$ in this case, and therefore no positive sorting fee can be found such that the rich would be willing to pay the fee and be in the rich group and the poor would not be willing to join.

**Proposition 14** *No cutoff $\hat{y} \in Y$ constitutes a biased sorting equilibrium in Case 4.*

**Proof.** The fact that in Case 4 we have $\bar{E}(\hat{y}^*) < \bar{E}_p(\hat{y}^*)$ and $\underline{E}_r(\hat{y}^*) > \underline{E}(\hat{y}^*)$ $\forall \hat{y}^* \in (0, y_{\max})$ combined with strict increasingness of $U$ implies that inequality (15) cannot be satisfied for any $\hat{y} \in Y$. ■

The opposite to Case 4 happens in Case 3: if the rich group overestimates the group difference for every cutoff, while the poor group underestimates it, any cutoff $\hat{y} \in Y$ is a biased sorting equilibrium.

**Proposition 15** *Any cutoff $\hat{y} \in Y$ constitutes a biased sorting equilibrium in Case 3.*[27]

**Proof.** Case 3 implies that $\bar{E}(\hat{y}^*) > \bar{E}_p(\hat{y}^*)$ and $\underline{E}_r(\hat{y}^*) < \underline{E}(\hat{y}^*)$ $\forall \hat{y}^* \in (0, y_{\max})$. Together with the fact that $U$ is strictly increasing in both arguments, this implies that inequality (15) holds for all $\hat{y} \in Y$. ■

If we require the equilibrium partition to satisfy consistency, the equilibrium cutoff must satisfy

$$U(\hat{y}^*, \bar{E}(\hat{y}^*)) - U(\hat{y}^*, \underline{E}_r(\hat{y}^*)) = U(\hat{y}^*, \bar{E}_p(\hat{y}^*)) - U(\hat{y}^*, \underline{E}(\hat{y}^*)) = b \tag{16}$$

(and therefore both differences must be strictly positive because $b > 0$).

**Proposition 16** *No cutoff $\hat{y} \in Y$ constitutes a biased sorting equilibrium with consistency in Case 3.*

**Proof.** Case 3 implies that $\bar{E}(\hat{y}^*) < \bar{E}_p(\hat{y}^*)$ and $\underline{E}_r(\hat{y}^*) > \underline{E}(\hat{y}^*)$ $\forall \hat{y}^* \in (0, y_{\max})$. If $U$ is strictly increasing everywhere, condition (16) cannot be satisfied for any $\hat{y} \in Y$. ■

Only Case 1 and Case 2 allow for the existence of an interior biased sorting equilibrium with consistency. In fact, I find that in those two cases an interior equilibrium always exists:

**Proposition 17** *A biased sorting equilibrium with consistency always exists in Case 1 and Case 2.*

**Proof.** Remember that the rich are correct at 0 and the poor at $y_{\max}$. Therefore, it holds that

$$U(0, \bar{E}(0)) - U(0, \underline{E}_r(0)) > (<) U(0, \bar{E}_p(0)) - U(0, \underline{E}(0))$$

[27] If I would be meticulously diligent, I would have to exclude $\hat{y} = 0$ as a potential equilibrium cutoff in this and all following Propositions. The reason is that strictly speaking $\hat{y} = 0$ cannot be an equilibrium cutoff due to my definition of the partition as $\{[0, \hat{y}), [\hat{y}, y_{\max}]\}$. $\hat{y} = 0$ would imply that $S_0$ would be empty, which is not possible because the empty set cannot be an element of a partition. Therefore, $\hat{y} = 0$ is technically not even included in my sorting equilibrium definition. If I had defined the partition the other way round, i.e. $\{[0, \hat{y}], (\hat{y}, y_{\max}]\}$, then the same would hold for $y_{\max}$.

and

$$U(y_{\max}, \bar{E}(y_{\max})) - U(y_{\max}, \underline{E}_r(y_{\max})) < (>) U(y_{\max}, \bar{E}_p(y_{\max})) - U(y_{\max}, \underline{E}(y_{\max}))$$

As both $U$ and the belief function are continuous, there must be an interior $\hat{y}^* \in (0, y_{\max})$ such that

$$U(\hat{y}^*, \bar{E}(\hat{y}^*)) - U(\hat{y}^*, \underline{E}_r(\hat{y}^*)) = U(\hat{y}^*, \bar{E}_p(\hat{y}^*)) - U(\hat{y}^*, \underline{E}(\hat{y}^*))$$

■

### A.2.1 Linear utility function

If $U$ is linear in both arguments, it is strictly speaking not in my set of analyzed utility functions, because it is not strictly increasing (and also not strictly supermodular) whenever one of the arguments is 0. However, for reasons of simplicity, I use a linear utility function for my analysis in Section 6 (and note that such a utility function satisfies all my requirements whenever both arguments are strictly positive). If $U(x, y) = xy$ the equilibrium condition translates to

$$\hat{y}^*[\bar{E} - \underline{E}_r] = \hat{y}^*[\bar{E}_p - \underline{E}]$$

It is immediate to see that this condition will always be satisfied at $\hat{y} = 0$ (also for Case 3 and Case 4). However, $\hat{y} = 0$ cannot be a biased sorting equilibrium cutoff according to my definition, because I require the sorting fee $b$ to be strictly positive, and hence $U(\hat{y}^*, \bar{E}(\hat{y}^*)) - U(\hat{y}^*, \underline{E}_r(\hat{y}^*))$ must be striclty positive in any equilibrium. Therefore, $\hat{y} = 0$ can never constitute a biased sorting equilibrium cutoff if $U$ is linear.

For reasons of completeness, I therefore restate Propositions 14 - 16 for a linear utility function (Proposition 17 doesn't change):

**Proposition 18** *No cutoff $\hat{y} \in Y$ constitutes a biased sorting equilibrium in Case 4.*

**Proof.** The fact that in Case 4 we have $\bar{E}(\hat{y}^*) < \bar{E}_p(\hat{y}^*)$ and $\underline{E}_r(\hat{y}^*) > \underline{E}(\hat{y}^*)$ $\forall \hat{y}^* \in (0, y_{\max})$ combined with strict increasingness of $U$ implies that inequality (15) cannot be satisfied for any $\hat{y} \in Y$. If $U$ is linear (and hence not strictly increasing at $\hat{y} = 0$) then inequality (15) is trivially satisfied for $\hat{y} = 0$, but $\hat{y} = 0$ cannot be a biased sorting equilibrium cutoff according to my definition, because I require the sorting fee $b$ to be strictly positive, and hence $U(\hat{y}^*, \bar{E}(\hat{y}^*)) - U(\hat{y}^*, \underline{E}_r(\hat{y}^*))$ must be striclty positive in any equilibrium. ■

**Proposition 19** *Any cutoff $\hat{y} \in (0, y_{\max}]$ constitutes a biased sorting equilibrium in Case 3.*

**Proof.** Case 3 implies that $\bar{E}(\hat{y}^*) > \bar{E}_p(\hat{y}^*)$ and $\underline{E}_r(\hat{y}^*) < \underline{E}(\hat{y}^*)$ $\forall \hat{y}^* \in (0, y_{\max})$. If $U$ is strictly increasing in both arguments, this implies that inequality (15) holds for all $\hat{y} \in Y$. If $U$ is linear (and hence not strictly increasing at $\hat{y} = 0$) then inequality (15) is trivially satisfied for $\hat{y} = 0$, but $\hat{y} = 0$ cannot be a biased sorting equilibrium cutoff. ■

**Proposition 20** *No cutoff $\hat{y} \in (0, y_{\max})$ constitutes a biased sorting equilibrium with consistency in Case 3.*

**Proof.** Case 3 implies that $\bar{E}(\hat{y}^*) < \bar{E}_p(\hat{y}^*)$ and $\underline{E}_r(\hat{y}^*) > \underline{E}(\hat{y}^*)$ $\forall \hat{y}^* \in (0, y_{\max})$. If $U$ is strictly increasing everywhere, condition (16) cannot be satisfied for any $\hat{y} \in Y$. If $U$ is linear then condition (16) is trivially satisfied for $\hat{y} = 0$, but $\hat{y} = 0$ cannot be a biased sorting equilibrium cutoff. ■

### A.2.2 Uniqueness: Linear utility function

For Case 1 and Case 2, the following sufficient conditions for uniqueness of an interior biased sorting equilibrium with consistency can be stated if $U$ is linear:

**Proposition 21** *If $U(x, y) = xy$, and people's misperceptions converge to the truth monotonically, i.e.*

$$\frac{d\left|\bar{E}(\hat{y}) - \bar{E}_p(\hat{y})\right|}{d\hat{y}} < 0 \quad \textit{and} \quad \frac{d\left|\underline{E}_r(\hat{y}) - \underline{E}(\hat{y})\right|}{d\hat{y}} > 0 \;\; \forall \hat{y} \in (0, y_{\max}) \qquad (17)$$

*there always exists a unique biased sorting equilibrium with consistency.*

**Proof.** The equilibrium cutoff must satisfy

$$\bar{E}(\hat{y}^*) - \underline{E}_r(\hat{y}^*) = \bar{E}_p(\hat{y}^*) - \underline{E}(\hat{y}^*)$$

Suppose we are in Case 1 (the argument can be made analogously for Case 2). Then the conditions in (17) become

$$\frac{d(\bar{E}(\hat{y}) - \bar{E}_p(\hat{y}))}{d\hat{y}} < 0 \quad \text{and} \quad \frac{d(\underline{E}_r(\hat{y}) - \underline{E}(\hat{y}))}{d\hat{y}} > 0 \;\; \forall \hat{y} \in (0, y_{\max})$$

This implies that the distance between the correct difference in average group incomes, $\underline{E}(\hat{y}) - \underline{E}(\hat{y})$, and the poor's perceived group difference, $\bar{E}_p(\hat{y}) - \underline{E}(\hat{y})$, which can be written as $\bar{E}(\hat{y}) - \bar{E}_p(\hat{y})$, is monotonically decreasing in $\hat{y}$, while the opposite holds for the distance between the correct group difference and the rich's perceived group difference (which can be written as $\underline{E}_r(\hat{y}) - \underline{E}(\hat{y})$). This means that there can be only one $\hat{y}$ for which the distance between the correct group differences and the group's perceived group differences is the same, and therefore (17) guarantees that the perceived benefits of sorting of the rich and of the poor only cut once in the interval $(0, y_{\max})$. ■

### A.2.3 Uniqueness in Case 1: General utility function

For a general utility function, we also need to impose conditions on the shape of the utility function to ensure uniqueness.

**Proposition 22** *If people are biased according to Case 1 and people's misperceptions converge to the truth monotonically, i.e.*

$$\frac{d\left|\bar{E}(\hat{y})-\bar{E}_p(\hat{y})\right|}{d\hat{y}}<0 \quad \textit{and} \quad \frac{d\left|\underline{E}_r(\hat{y})-\underline{E}(\hat{y})\right|}{d\hat{y}}>0 \quad \forall\hat{y}\in(0,y_{\max})$$

*and additionally it holds that at any $\hat{y}^*$ for which*

$$U(\hat{y}^*,\bar{E}(\hat{y}^*))-U(\hat{y}^*,\underline{E}_r(\hat{y}^*))=U(\hat{y}^*,\bar{E}_p(\hat{y}^*))-U(\hat{y}^*,\underline{E}(\hat{y}^*))$$

*holds we have that*

$$U_1(\hat{y}^*,\bar{E}(\hat{y}^*))-U_1(\hat{y}^*,\underline{E}_r(\hat{y}^*))\leq U_1(\hat{y}^*,\bar{E}_p(\hat{y}^*))-U_1(\hat{y}^*,\underline{E}(\hat{y}^*))$$

$$U_2(\hat{y}^*,\underline{E}_r(\hat{y}^*))\geq U_2(\hat{y}^*,\underline{E}(\hat{y}^*))$$

*and*

$$U_2(\hat{y}^*,\bar{E}(\hat{y}^*))\leq U_2(\hat{y}^*,\bar{E}_p(\hat{y}^*))$$

*there always exists a unique biased sorting equilibrium with consistency.*

**Proof.** At any equilibrium cutoff $\hat{y}^*$ such that

$$U(\hat{y}^*,\bar{E}(\hat{y}^*))-U(\hat{y}^*,\underline{E}_r(\hat{y}^*))=U(\hat{y}^*,\bar{E}_p(\hat{y}^*))-U(\hat{y}^*,\underline{E}(\hat{y}^*))$$

these conditions ensure that the slope of the function on the LHS is smaller than the slope of the function on the RHS, which implies that the two functions can only intersect once. ■

**Remark 3** *The conditions in Proposition 22 are sufficient conditions for uniqueness, because they ensure that the slope of the left hand side of the equilibrium condition is strictly smaller than the slope of the right hand side at any intersection.*

**Remark 4** *Similar sufficient conditions can be found for Case 2 type misperceptions, where both groups overestimate group differences.*

#### A.2.4 Conditions for a unique equilibrium above the median with linear utility

**Proposition 23** *If $U$ is linear and people are biased according to Case 1, sufficient conditions for a unique equilibrium cutoff $\hat{y}^*$ above median are*

$$\bar{E}(\hat{y})-\bar{E}_p(\hat{y}) \textit{ monotonically decreasing in } \hat{y}$$

*and*

$$\underline{E}_r(\hat{y})-\underline{E}(\hat{y}) \textit{ monotonically increasing in } \hat{y}$$

*and additionally,*

$$\bar{E}_p(y^M)+\underline{E}_r(y^M)<2E$$

**Proof.** The first two conditions guarantee uniqueness. Concerning the last condition, note that if $\bar{E} - \bar{E}_p$ is monotonically increasing and $\underline{E}_r - \underline{E}$ is monotonically decreasing in $\hat{y}$, then

$$\bar{E}_p - \underline{E} < \bar{E} - \underline{E}_r$$

for all $\hat{y}$ below the unique equilibrium cutoff, and the inequality must hold in the other direction above the unique equilibrium cutoff. That implies

$$\bar{E}_p + \underline{E}_r < \bar{E} + \underline{E}$$

for all $\hat{y}$ below the equilibrium cutoff, and

$$\bar{E}_p + \underline{E}_r > \bar{E} + \underline{E}$$

for all $\hat{y}$ above the equilibrium cutoff. If the equilibrium should lie above the median, then at the median it must be that

$$\bar{E}_p(y^M) + \underline{E}_r(y^M) < \bar{E} + \underline{E},$$

because the median must be below the cutoff. The fact that

$$E = (1-F)\bar{E} + F\underline{E} = \frac{\bar{E} + \underline{E}}{2}$$

at the median proves the claim. ∎

#### A.2.5 Examination of the unique binary biased sorting equilibrium

As established in Section 6, the unique equilibrium cutoff is characterized by

$$\hat{y}^* = \frac{a(1-F)\bar{E} + F\underline{E}}{a(1-F) + F}$$

and hence it is the fixed point of

$$h(\hat{y}) = \frac{a(1-F)\bar{E} + F\underline{E}}{a(1-F) + F}$$

Therefore, the equilibrium cutoff is exactly where the 45 degree line cuts the function $h$. To ensure that there can only be one such intersection point, I can calculate

$$h'(\hat{y}) = \frac{\left[\begin{array}{c} \left(-af\bar{E} + a(1-F)\frac{\partial \bar{E}}{\partial \hat{y}} + f\underline{E} + F\frac{\partial \underline{E}}{\partial \hat{y}}\right)(a(1-F)+F) \\ -\left(a(1-F)\bar{E} + F\underline{E}\right)(-af+f) \end{array}\right]}{(a(1-F)+F)^2}$$

which can be simplified to

$$h'(\hat{y}) = \frac{(1-a)f}{(a(1-F)+F)^2}\left[a(1-F)(\hat{y} - \bar{E}) + F(\hat{y} - \underline{E})\right]$$

This implies that $h$ has a local extremum $\hat{y}^{**}$ characterized by

$$a(1-F)(\hat{y}-\bar{E})+F(\hat{y}-\underline{E})=0$$

or equivalently

$$\hat{y}^{**}=\frac{a(1-F)\bar{E}+F\underline{E}}{a(1-F)+F} \tag{18}$$

This is exactly the equation that characterizes the equilibrium cutoff and the fixed point of $h$. Hence whenever the 45 degree line cuts $h$ it must be at a local extremum, i.e. where the slope of $h$ is 0. This means that at any intersection, the 45 degree line cuts $h$ from below, which implies that such an intersection can only happen once. It follows that $h$ will have a unique fixed point and the equilibrium cutoff is unique.

The local extremum of $h$ characterized by (18) is a local maximum if $a>1$ and a local minimum if $a<1$. This can be seen from noting that

$$\begin{aligned}h''(\hat{y})=&\frac{(1-a)f'}{(a(1-F)+F)^2}\left[a(1-F)(\hat{y}-\bar{E})+F(\hat{y}-\underline{E})\right]\\&+\frac{(1-a)f}{(a(1-F)+F)}\\&-\frac{2(1-a)^2f^2\left[a(1-F)(\hat{y}-\bar{E})+F(\hat{y}-\underline{E})\right]}{(a(1-F)+F)^3}\end{aligned}$$

At $\hat{y}^{**}$ we know that

$$a(1-F)(\hat{y}-\bar{E})+F(\hat{y}-\underline{E})=0$$

and thus the first and the third term drop out of the second derivative and we get

$$h''(\hat{y}^{**})=\frac{(1-a)f}{(a(1-F)+F)}$$

Therefore, $\hat{y}^{**}$ is a local maximum if $a>1$ and a local minimum at $a<1$. Figures 4 and 3 depict the intersection of $h$ and the 45 degree line for $a<1$ and $a>1$. If $a=1$ the problem becomes very simple, as the expression for $h$ reduces to

$$h(\hat{y})=E,$$

i.e. $h$ is just a horizontal straight line at $E$ and the unique equilibrium cutoff is at $E$.

## A.3 Case 1: How $\hat{y}$ changes with naivety

As noted above, the equilibrium cutoff depends on the naivety of the rich and the poor via a single parameter, $\frac{\beta}{\gamma}=a$, which describes the severity of the

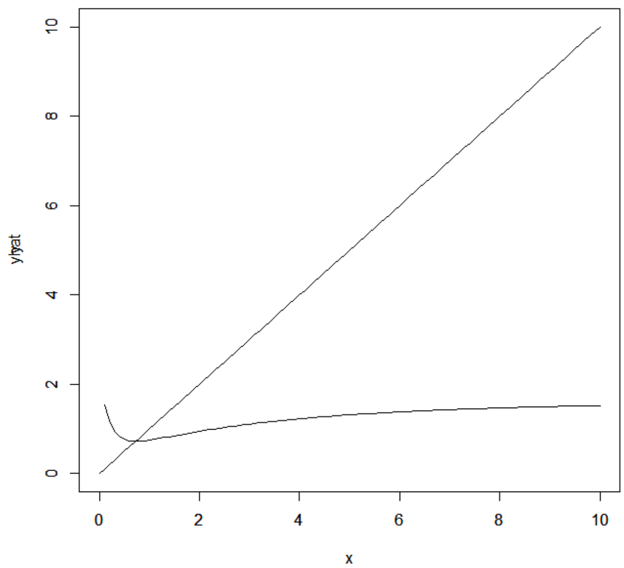


Figure 3: Equilibrium cutoff $\hat{y}^*$ if $a < 1$

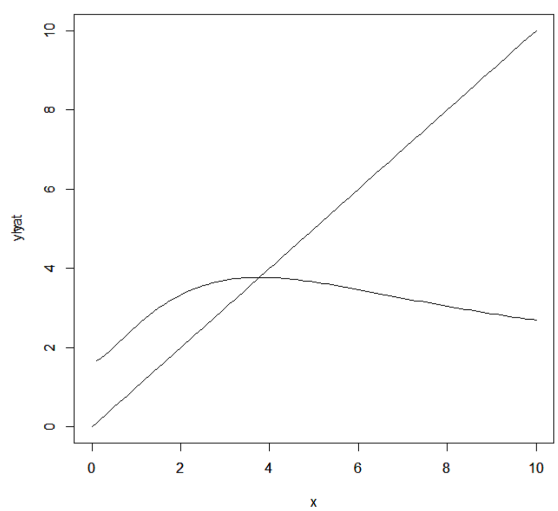


Figure 4: Equilibrium cutoff $\hat{y}^*$ if $a > 1$

poor's naivety relative to the rich's. If $a = 1$ then both groups are "equally naive", if $a > 1$ then the poor are more naive than the rich. Writing

$$\hat{y}^* = \frac{a(1-F)\bar{E} + F\underline{E}}{a(1-F)+F} \tag{19}$$

I can investigate how $\hat{y}^*$ changes with $a$:

$$\begin{aligned}
(1-F)\bar{E}da + (-af\bar{E} + a(1-F)\frac{\bar{E}-\hat{y}^*}{1-F}f + f\underline{E} + F\frac{(\hat{y}^*-\underline{E})}{F}f)d\hat{y}^* \\
= (a(1-F) + F + \hat{y}^*(-af+f))d\hat{y}^* + (1-F)\hat{y}^*da \\
(1-F)(\bar{E}-\hat{y}^*)da = [af\bar{E} - a(\bar{E}-\hat{y}^*)f - \underline{E}f - (\hat{y}^*-\underline{E})f \\
+a(1-F) + F + \hat{y}^* f(1-a)]d\hat{y}^* \\
\frac{(1-F)(\bar{E}-\hat{y}^*)}{a(1-F)+F} = \frac{d\hat{y}^*}{da} > 0
\end{aligned} \tag{20}$$

(20) $\hat{y}^*$ is increasing in the degree of naivety of the poor relative to the rich. The higher $a$, the more the poor tend to underestimate the benefits of sorting (relative to the rich) and hence the more they need to see of the whole distribution relative to the rich to have the same perceived benefits of sorting as the rich.

## A.4 A median-preserving spread on the lognormal distribution and monopolist profits

Recall that the monopolist's profits from offering one cutoff (which in equilibrium will be at $E$ if $a = 1$) can be written as

$$E(E-\underline{E})[1-\gamma F(E)(1-F(E))] - c$$

For the lognormal distribution, this becomes

$$\Pi = E\left[E\left(1 - \frac{\Phi\left(\frac{\ln\hat{y}-\mu}{\sigma} - \sigma\right)}{\Phi\left(\frac{\ln\hat{y}-\mu}{\sigma}\right)}\right)\left(1-\gamma\Phi\left(\frac{\ln\hat{y}-\mu}{\sigma}\right) + \gamma\Phi\left(\frac{\ln\hat{y}-\mu}{\sigma}\right)\right)\right] - c$$

which can be simplified to

$$\begin{aligned}
\Pi &= E\left[E\left(1-\frac{1-\Phi\left(\frac{\sigma}{2}\right)}{\Phi\left(\frac{\sigma}{2}\right)}\right)\left(1-\gamma\Phi\left(\frac{\sigma}{2}\right)+\gamma\left[\Phi\left(\frac{\sigma}{2}\right)\right]^2\right)\right] - c \\
&= E^2\left(\frac{2\Phi\left(\frac{\sigma}{2}\right)-1}{\Phi\left(\frac{\sigma}{2}\right)}\right)\left(1-\gamma\Phi\left(\frac{\sigma}{2}\right)+\gamma\left[\Phi\left(\frac{\sigma}{2}\right)\right]^2\right) - c
\end{aligned}$$

because $\ln\hat{y} = \mu + \sigma^2$ if $\hat{y} = E$.

I find that

$$\frac{d\Pi}{d\sigma} = 2\sigma E^2 \left(\frac{2\Phi\left(\frac{\sigma}{2}\right)-1}{\Phi\left(\frac{\sigma}{2}\right)}\right)\left(1-\gamma\Phi\left(\frac{\sigma}{2}\right)+\gamma\left[\Phi\left(\frac{\sigma}{2}\right)\right]^2\right)$$
$$+E^2\left[\frac{\phi(\frac{\sigma}{2})\frac{1}{2}}{\Phi\left(\frac{\sigma}{2}\right)^2}\right]\left(1-\gamma\Phi\left(\frac{\sigma}{2}\right)+\gamma\left[\Phi\left(\frac{\sigma}{2}\right)\right]^2\right)$$
$$+E^2\left(\frac{2\Phi\left(\frac{\sigma}{2}\right)-1}{\Phi\left(\frac{\sigma}{2}\right)}\right)\gamma\phi\left(\frac{\sigma}{2}\right)\left(\Phi\left(\frac{\sigma}{2}\right)-\frac{1}{2}\right)$$

As $\Phi\left(\frac{\sigma}{2}\right) > \frac{1}{2}$ all of the terms are positive and hence the monopolist's profit always increases if $\sigma$ increases.

**Proposition 24** *If income is lognormally distributed, an increase in inequality in the form of a median-preserving spread increases the monopolist's revenues from offering the sorting technology.*

## A.5 Sufficient conditions for Assumption 2

$$\frac{y^M}{E_p} \le \frac{E}{E_r}$$
$$\iff y^M(F\underline{E}_r + (1-F)\bar{E}) \le E(F\underline{E} + (1-F)\bar{E}_p)$$

If $\beta = \gamma$, this can be simplified to

$$\beta(y^M F^2(E-\underline{E}) + E(1-F)^2(\bar{E}-E)) \le E(E-y^M)$$

Noting that

$$E - \underline{E} = (1-F)(\bar{E}-\underline{E})$$

and

$$\bar{E} - E = F(\bar{E}-\underline{E})$$

I can further simplify to

$$\beta F(1-F)\left(F\frac{y^M}{E} + (1-F)\right)(\bar{E}-\underline{E}) \le (E-y^M)$$

Given that $F(1-F) < 0.25$ (because $y^M < E$) and $\frac{y^M}{E} < 1$, I have that

$$\beta F(1-F)\left(F\frac{y^M}{E} + (1-F)\right)(\bar{E}-\underline{E}) < \beta\frac{(\bar{E}-\underline{E})}{4}$$

and it follows that

$$\beta\frac{(\bar{E}-\underline{E})}{4} \le E - y^M$$

is a sufficient condition for

$$\frac{y^M}{E_p} \le \frac{E}{E_r}.$$

## A.6 Detailed calculations for Section 6.4

Average income $E$ does not change due to a mean-preserving spread and hence[28]

$$\Delta E = F\Delta \underline{E} + (1-F)\Delta \bar{E} = 0, \tag{21}$$

Average perceived income of the poor, $E_p$, declines, because

$$\Delta E_p = F\Delta \underline{E} + (1-F)\Delta \bar{E}_p$$

and

$$\bar{E}_p = \beta(1-F)\hat{y} + (1-\beta(1-F))\bar{E}$$

which implies

$$\Delta \bar{E}_p = (1-\beta(1-F))\Delta \bar{E} < \Delta \bar{E} \tag{22}$$

(as $\hat{y} = E$ doesn't change). The change in $\frac{y^M}{E_p}$ amounts to

$$\Delta \left(\frac{y^M}{E_p}\right) = \frac{\Delta y^M E_p - y^M \Delta E_p}{(E_p)^2} = \left(\frac{\Delta y^M}{y^M} - \frac{\Delta E_p}{E_p}\right)\frac{y^M}{E_p}$$

and thus the percentage change in $\frac{y^M}{E_p}$ is $\frac{\Delta y^M}{y^M} - \frac{\Delta E_p}{E_p}$, which is smaller (in absolute terms) than the percentage change of $\frac{y^M}{E}$ in the unbiased case, because $\frac{\Delta E_p}{E_p} < 0$. In the following I show that if $\left|\frac{\Delta E_p}{E_p}\right|$ is large enough relative to $\left|\frac{\Delta y^M}{y^M}\right|$, the median earner will even think that inequality has decreased, i.e. the percentage change in $\frac{y^M}{E_p}$ (and hence also the absolute change in $\frac{y^M}{E_p}$) can be positive:

From (21) and (22) it follows that

$$\Delta E_p = -(1-F)\Delta \bar{E} + (1-F)\Delta \bar{E}_p = -\beta(1-F)^2 \Delta \bar{E}$$

Furthermore,

$$E_p = F\underline{E} + (1-F)\bar{E}_p = E - \beta(1-F)^2(\bar{E} - E)$$

and therefore

$$\frac{\Delta E_p}{E_p} = \frac{-\beta(1-F)^2 \Delta \bar{E}}{E - \beta(1-F)^2(\bar{E}-E)} = \frac{\beta(1-F)F\Delta \underline{E}}{E - \beta(1-F)^2(\bar{E}-E)}$$

(using (21) again). Hence, I get

$$\frac{\Delta y^M}{y^M} - \frac{\Delta E_p}{E_p} > 0 \iff \frac{\Delta y^M}{y^M} > \frac{\beta(1-F)F\Delta \underline{E}}{E - \beta(1-F)^2(\bar{E}-E)}$$

[28] And note that I require the mean-preserving spread to be such that $F(\hat{y}^*) = F(E)$ doesn't change.

$$\iff \frac{\Delta y^M}{\Delta \underline{E}} < \frac{\beta y^M (1-F)F}{E - \beta(1-F)^2(\bar{E}-E)} \tag{23}$$

(where both sides are positive).For a given $\frac{\Delta y^M}{\Delta \underline{E}}$ this condition is more likely to be satisfied if $\beta$ is large, because

$$\frac{\partial}{\partial \beta}\left(\frac{\beta}{E - \beta(1-F)^2(\bar{E}-E)}\right) = \frac{E}{\left[E - \beta(1-F)^2(\bar{E}-E)\right]^2} > 0$$

and hence the RHS is increasing in $\beta$. Furthermore, $(1-F(E))F(E)$ should not be too small, i.e. the income distribution cannot be too positively skewed, such that $F(E)$ is not too far above 0.5. Note however, that such a mean-preserving spread can be constructed for any given income distribution, by ensuring that $\Delta y^M$ and $\Delta \underline{E}$ are such that (23) holds (the right hand side is determined by parameters of the model, and the mean-preserving spread has to be designed such that the ratio on the LHS satisfies the constraint). Hence, I can conclude that

**Lemma 4** *For any $\beta > 0$ there exists a mean-preserving spread of the income distribution such that an increase in inequality leads to a decrease in the median earner's perceived degree of inequality.*

Now let me examine the absolute change of $\frac{y^M}{E}$ and $\frac{y^M}{E_p}$: I want to derive sufficient conditions for the absolute decrease in perceived equality to be smaller under segregation, i.e.

$$\Delta\left(\frac{y^M}{E}\right) < \Delta\left(\frac{y^M}{E_p}\right) \tag{24}$$

(because both sides of this inequality are negative). Lemma 4 shows that I can always construct a mean-preserving spread satisfying (23) such that perceived equality $\frac{y^M}{E_p}$ increases under segregation (in which case inequality (24) trivially holds, because $\frac{y^M}{E}$ will always decrease). However, less strong conditions can be derived in order for (24) to hold without perceived inequality having to decrease:

$$\Delta\left(\frac{y^M}{E}\right) < \Delta\left(\frac{y^M}{E_p}\right) \iff$$

$$\frac{\Delta y^M}{E} < \frac{\Delta y^M E_p - y^M \Delta E_p}{(E_p)^2} = \frac{\Delta y^M}{E_p} - \frac{y^M \Delta E_p}{(E_p)^2}$$

$$\iff \frac{\Delta y^M}{E} < \frac{\Delta y^M}{E - \beta(1-F)^2(\bar{E}-E)} - \frac{y^M \beta(1-F)^2 \Delta \bar{E}}{\left(E - \beta(1-F)^2(\bar{E}-E)\right)^2}$$

$$\iff \Delta y^M \left(\frac{1}{E} - \frac{1}{E - \beta(1-F)^2(\bar{E}-E)}\right) < \frac{y^M \beta(1-F)^2 \Delta \bar{E}}{\left(E - \beta(1-F)^2(\bar{E}-E)\right)^2}$$

$$\iff \frac{\Delta y^M}{y^M}\left(\frac{-\beta(1-F)^2(\bar{E}-E)}{E(E - \beta(1-F)^2(\bar{E}-E))}\right) < \frac{\beta(1-F)^2 \Delta \bar{E}}{\left(E - \beta(1-F)^2(\bar{E}-E)\right)^2}$$

$$\iff \frac{\Delta y^M}{y^M}\left(\frac{-(\bar{E}-E)}{E}\right) < \frac{\Delta \bar{E}}{E-\beta(1-F)^2(\bar{E}-E)}$$

$$\iff \frac{\Delta y^M}{y^M}\left(\frac{-F(\bar{E}-\underline{E})}{E}\right) < \frac{\frac{-F\Delta \underline{E}}{1-F}}{E-\beta(1-F)^2(\bar{E}-E)}$$

$$\iff -\frac{\Delta y^M}{y^M}\left(\frac{(1-F)(\bar{E}-\underline{E})}{E}\right) < \frac{-\Delta \underline{E}}{E-\beta(1-F)^2(\bar{E}-E)}$$

$$\iff \frac{\Delta y^M}{\Delta \underline{E}}\left(\frac{E-\underline{E}}{E}\right) < \frac{y^M}{E-\beta(1-F)^2(\bar{E}-E)}$$

$$\iff \frac{\Delta y^M}{\Delta \underline{E}} < \frac{y^M E}{\left(E-\beta(1-F)^2(\bar{E}-E)\right)(E-\underline{E})} = \frac{y^M E}{E_p(E-\underline{E})}$$

For a given mean-preserving spread, this inequality is more likely to hold if $\beta$ is large (such that $E_p$ is small relative to $E$). Note however, that it is always possible to construct a mean-preserving spread that satisfies this inequality, by designing $\Delta y^M$ and $\Delta \underline{E}$ accordingly.

**Lemma 5** *The (absolute) decrease in $\frac{y^M}{E_p}$ is smaller than the (absolute) decrease in $\frac{y^M}{E}$ iff the mean-preserving spread is such that*

$$\frac{\Delta y^M}{\Delta \underline{E}} < \frac{y^M E}{\left(E-\beta(1-F)^2(\bar{E}-E)\right)(E-\underline{E})} \tag{25}$$

In the absence of segregation, the change in the median earner's preferred tax rate due to a mean-preserving spread is given by[29]

$$\Delta t^* = \tau'^{-1}\left(\frac{\dot{y}^M}{E}\right) - \tau'^{-1}\left(\frac{y^M}{E}\right)$$

If society is segregated, the change in the median earner's preferred tax rate amounts to[30]

$$\Delta \tilde{t}^* = \tau'^{-1}\left(\frac{\dot{y}^M}{\dot{E}_p}\right) - \tau'^{-1}\left(\frac{y^M}{E_p}\right)$$

If the conditions of Lemma (5) hold, the decrease in $\frac{y^M}{E_p}$ is smaller than the decrease in $\frac{y^M}{E}$. Furthermore, I know that $\frac{y^M}{E_p} > \frac{y^M}{E}$. Together with the fact that $\tau''(t) \leq 0$ and $\tau'''(t) \geq 0$, which implies that $\tau'^{-1}$ is decreasing and convex, this gives

$$\Delta \tilde{t}^* < \Delta t^*$$

**Lemma 6** *If the mean-preserving spread is such that (25) holds, the increase in the preferred tax rate is less in a segregated society than in the absence of segregation.*

[29] Notation: $\dot{y}^M$ denotes median income after the mean-preserving spread.

[30] Notation: $\dot{y}^M$ denotes median income after the mean-preserving spread and $\dot{E}_p$ denotes the poor group's perception of average income after the mean-preserving spread.

## A.7 Detailed calculations for Section 6.5

If a mean-preserving spread leads to economic segregation, the median earner's demand for redistribution declines if

$$\frac{y^M}{E} < \frac{\dot{y}^M}{E_p}$$

$$\iff y^M(E - \beta(1-F)^2(\bar{E} + \Delta\bar{E} - E) < y^M E + E\Delta y^M$$

$$\iff \frac{\beta(1-F)^2(\bar{E} + \Delta\bar{E} - E)}{E} > -\frac{\Delta y^M}{y^M} = \left|\frac{\Delta y^M}{y^M}\right| \tag{26}$$

For a given mean-preserving spread this inequality holds if $\beta$ is large enough (i.e. people are sufficiently naive) and the increase in average income in the rich group is large enough relative to the decline in median income. Again, it is immediate to see that a mean-preserving spread satisfying (26) can always be constructed by designing $\Delta\bar{E}$ and $\Delta y^M$ accordingly.

## A.8 The effect of general changes in the shape of the income distribution on the demand for redistribution if society is segregated

What happens to people's preferred redistribution rate if inequality between groups changes when people are already segregated?

Recall that the equilibrium cutoff is given by

$$\hat{y}^* = \frac{a(1-F)\bar{E} + F\underline{E}}{a(1-F) + F} \tag{27}$$

i.e. $\hat{y}^*$ is the fixed point of the function

$$h(\hat{y}) = \frac{a(1-F)\bar{E} + F\underline{E}}{a(1-F) + F}$$

As described in Section 6, (27) has a unique fixed point. If $a = 1$, this fixed point is at average income $E$. For $a > 1$ the intersection between $h(\hat{y})$ and the 45 degree line looks like Figure (6), if $a < 1$ then it looks like Figure (5) (and if $a = 1$, $\hat{y}^*$ is where the 45 degree line intersects with the horizontal line at $E$). From these graphs it is immediate to see that the impact of an increase in inequality on the equilibrium cutoff depends on how this increase in inequality affects $h(\hat{y})$ (and thus the intersection of the 45 degree line with $h(\hat{y})$).

Now clearly if $\underline{E}$ goes down while everything else stays the same, $h(\hat{y})$ just shifts down, and the intersection with the 45 degree line (=the equilibrium cutoff $\hat{y}^*$) goes down (both if $a > 1$ and if $a < 1$) Hence, the new equilibrium cutoff will be lower. The opposite happens if $\bar{E}$ goes up, i.e. if the rich group gets richer on average, again ceteris paribus: Then it is straightforward to see from (27) that the new equilibrium cutoff will be higher.

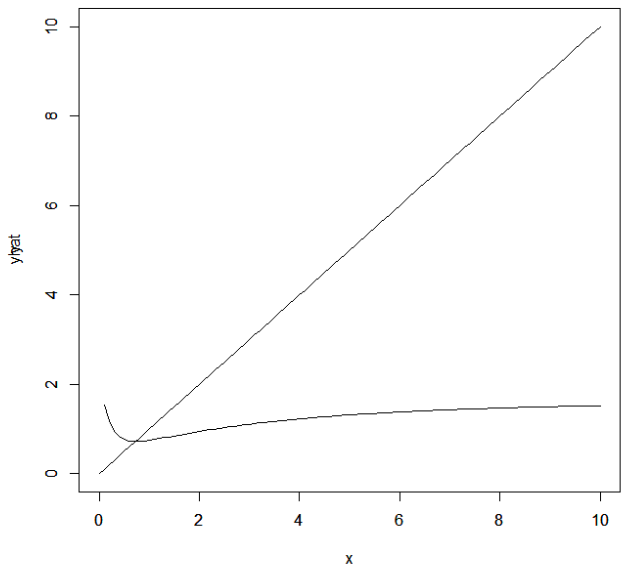


Figure 5: Equilibrium cutoff $\hat{y}^*$ if $a < 1$

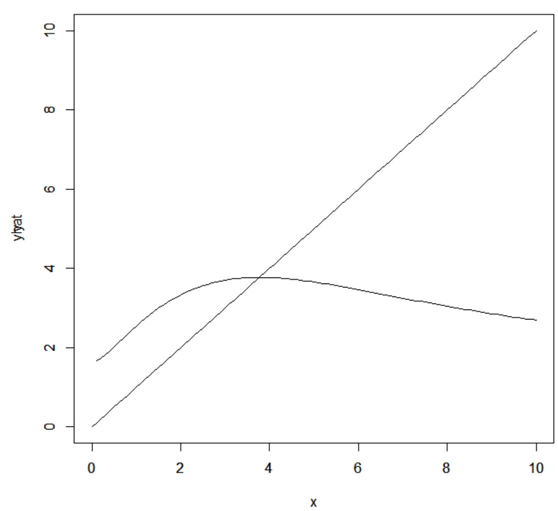


Figure 6: Equilibrium cutoff $\hat{y}^*$ if $a > 1$

Suppose that both things happen, so $\bar{E}$ increases, while $\underline{E}$ decreases (while $F$ and $1-F$ stay constant). Then whether the new equilibrium cutoff is higher or lower than the old one depends on $a$ and $F$ (resp. $1-F$) : if $a$ is high, or $1-F$ is high, such that $a(1-F)\Delta\bar{E}+F\Delta\underline{E}>0$, then the new equilibrium cutoff will be higher, if $a$ and/or $1-F$ is low, then the new equilibrium cutoff will be lower. If $a=1$ (meaning both groups are equally naive) then the cutoff is always $E$ and hence will go down if $E$ decreases due to this increase in inequality. $E$ decreases if $F(E)$ is high and $(1-F(E))$ is low, a feature that characterizes unequal distributions with positive skew.

If $\underline{E}$ decreases by $\Delta\underline{E}$ , then as I have shown above, $\hat{y}^*$ will go down. What happens to preferences for redistribution depends on the position of $y^M$: If $a=1$, the equilibrium cutoff is always at $E$, hence $y^M<\hat{y}^*=E$ before and after the decline in $\underline{E}$. If $y^M$ is sufficiently below the cutoff, such that preferences for redistribution do not overlap, the median earner is the decisive voter both before and after the change in $\underline{E}$[31]. Under these circumstances, a decrease in $\underline{E}$ and subsequently in $\hat{y}^*$ will mean that $\underline{E}$ decreases by $\Delta\underline{E}+\frac{\partial\underline{E}}{\partial\hat{y}}d\hat{y}^*$, $\bar{E}$ decreases by $\frac{\partial\bar{E}}{\partial\hat{y}}d\hat{y}^*$ and $E$ decreases by $F\Delta\underline{E}$ (the decreases in $\underline{E}$ and $\bar{E}$ due to the decrease in $\hat{y}^*$ cancel out with changes in $F$ and $1-F$ and do not affect $E$: clearly, where the cutoff is has no implications for average income). The decrease in $\underline{E}$ and subsequent fall in $\hat{y}^*$ will lead to decrease in $\bar{E}_p$ that is even higher than the decrease in $\bar{E}$, because as $\hat{y}^*$ decreases, the poor become more biased, i.e. $\bar{E}-\bar{E}_p$ increases. As $\bar{E}_p$ decreases, clearly also $E_p$ decreases, and if the poor are sufficiently biased, then this can lead to a situation where $\frac{y^M}{E_p}$ does not decrease, analogous to the unbiased case, but instead increases, because $E_p$ decreases by more than $y^M$:

$$d\left(\frac{y^M}{E_p}\right)=\frac{dy^M E_p-y^M dE_p}{E_p}$$

I find that

$$dE_p=F\Delta\underline{E}+\left[(1-F)^2\beta+f\beta(1-F)(\bar{E}-\hat{y})\right]d\hat{y}$$

and hence

$$|dE_p|>|F\Delta\underline{E}|=|dE|$$

This means that I can have

$$dy^M E_p-y^M dE_p=dy^M E_p-y^M dE_p>0$$

while

$$dy^M E-y^M dE<0$$

This implies that if $\underline{E}$ decreases ceteris paribus - and hence inequality increases - $\frac{y^M}{E_p}$ can increase instead of decrease (as $\frac{y^M}{E}$ would do in the unbiased case).

[31] By "preferences for redistribution do not overlap" I mean that the median earner should be sufficiently far away from the cutoff such that the person in the rich group with income just at the cutoff wants lower redistribution than the median earner.

A change in inequality that leads to a decrease of $\frac{y^M}{E}$ if people are unbiased, can lead to an increase in $\frac{y^M}{E_p}$ in the biased case. Hence, the new preferred tax rate after this increase in inequality can be lower than before. It is through the change in group composition, which affects people's bias, that an increase in inequality can lead to a decrease in the demand for redistribution.

Suppose that instead of $\underline{E}$ decreasing, $\bar{E}$ increases by $\Delta\bar{E}$ (which is also an increase in inequality between groups). Then the above analysis yields that $\hat{y}^*$ must increase. Hence, an increase in $\bar{E}$ means that the new equilibrium cutoff of the biased sorting equilibrium has to be higher. This means that $\bar{E}$ increases both because of the change in $\hat{y}^*$ and the shift $\Delta\bar{E}$ and also $\underline{E}$ increases due to the change in the cutoff. Hence, I have

$$dE_p = \left[f\underline{E} + F\frac{\partial\underline{E}}{\partial\hat{y}}\right]d\hat{y}^* + (1-F)d\bar{E}_p - f\bar{E}_p d\hat{y}^*$$

$$= (1-F)\Delta\bar{E} - \beta(1-F)^2\Delta\bar{E} + \left[\beta(1-F)^2 + f\beta(1-F)(\bar{E}-\hat{y})\right]d\hat{y}$$

The increase in $E_p$ will be smaller than the increase in $E$ (which is just $(1-F)\Delta\bar{E}$) if

$$\left[(1-F) + f(\bar{E}-\hat{y})\right]d\hat{y}^* - (1-F)\Delta\bar{E} < 0$$

or equivalently

$$\left[(1-F) + f(\bar{E}-\hat{y})\right]d\hat{y}^* < (1-F)\Delta\bar{E}$$

A necessary condition for this to hold is $\frac{d\hat{y}^*}{\Delta\bar{E}} < 1$, which can be shown to be satisfied. If $a = 1$ then $y^M < \hat{y}^* = E$ both before and after the change in $\bar{E}$, and assuming that $E$ is far enough above $y^M$ for the median earner to be decisive, $\frac{y^M}{E_p}$ will decrease by less than the corresponding fractions in the unbiased case, and the demand for redistribution will increase by less than in the unbiased case. However, we will clearly see an increase in the demand for redistribution if $\bar{E}$ increases, even if it is smaller than in the unbiased case.

If $\underline{E}$ decreases and $\bar{E}$ increases at the same time, in such a way that $a(1-F)\bar{E} + F\underline{E}$ decreases[32], $\hat{y}^*$ goes down. The change in $\frac{y^M}{E_p}$ amounts to

$$d\left(\frac{y^M}{E_p}\right) = \frac{dy^M E_p - y^M dE_p}{E_p}$$

where

$$E_p = F\underline{E} + (1-F)\bar{E}_p$$

[32]Note that if $a < 1$ (i.e. the poor are less naive than the rich) I can have that $a(1-F)\bar{E} + F\underline{E}$ decreases, while $E = (1-F)\bar{E} + F\underline{E}$ stays constant. An increase in inequality while $E$ stays constant is probably the closest to reality that this model can get, as I have not modelled growth here. If I would have modelled growth, then this increase in inequality where $\underline{E}$ decreases and $\bar{E}$ increases while $E$ stays constant would translate to $\underline{E}$ constant and $\bar{E}$ increasing while $E$ increases, which is probably what has happened over the last 30 years in the US and Europe. I have refrained from modelling growth here, because this would just have complicated the analysis ($\hat{y}^*$ would have to have a time trend etc.) while not changing the results about existence, uniqueness etc.

and

$$dE_p = F\Delta\underline{E} + \left(f\underline{E} + F\frac{\partial\underline{E}}{\partial\hat{y}}\right)d\hat{y} + (1-F)d\bar{E}_p - f\bar{E}_p d\hat{y}$$

$$= F\Delta\underline{E}+(1-F)(1-\beta(1-F))\Delta\bar{E}+\left[(1-F)^2\beta + f\beta(1-F)(\bar{E}-\hat{y})\right]d\hat{y}^* \quad (28)$$

Assuming that $a$ and $F(\hat{y})$ is such that $\hat{y}^*$ decreases due to an increase in $\bar{E}$ and a decrease in $\underline{E}$, it is clear from the last equation that $E_p$ can increase by less than $E$, and that $d\left(\frac{y^M}{E_p}\right)$ can be positive, and hence the demand for redistribution goes down as inequality increases.

If $a = 1$, (28) can be rewritten as

$$dE_p = (F\Delta\underline{E}+(1-F)\Delta\bar{E})\left[1+(1-F)^2\beta + f\beta(1-F)(\bar{E}-E)\right]-\beta(1-F)^2\Delta\bar{E} \quad (29)$$

using the fact that $\hat{y}^* = E$ in this case and

$$d\hat{y}^* = (1-F)\Delta\bar{E} + F\Delta\underline{E}$$

Clearly, in this case $\hat{y}^*$ decreases iff average income $E$ decreases due to the increase in inequality. Suppose that $\Delta\bar{E} = -\Delta\underline{E}$, then average income decreases iff $F(E) > (1-F(E))$, i.e. if the income distribution is positively skewed (which is what I assume in this paper). From (29), it follows that $dE_p < 0$ will hold if $(1-F)\Delta\bar{E} + F\Delta\underline{E} < 0$ (so as long as $d\hat{y}^* < 0$).[33] Hence, $d\left(\frac{y^M}{E_p}\right)$, which can be rewritten as

$$d\left(\frac{y^M}{E_p}\right) = \frac{dy^M}{E_p} - \frac{y^M}{E_p}\frac{dE_p}{E_p}$$

will be positive as long as $\frac{y^M}{E_p}$ is close to 1 and $\left(-\frac{dE_p}{E_p}\right) > \left(-\frac{dy^M}{E_p}\right)$. It is clear that a necessary condition for this is that $E_p$ decreases more than $y^M$, which is possible in the model,[34] because $E_p$ changes for two reasons: because actual average income $E$ decreases and because the poor's misperception (and hence underestimation of average income in the rich group) increases.

**Conclusion 7** *The effect of increasing inequality on support for redistribution if society is already segregated depends on the nature of the increase in inequality and on the rich and the poor's relative degree of naivety (resp. on a).*

- *If $a = 1$ and $\underline{E}$ decreases ceteris paribus, then the equilibrium cutoff will go down. This leads to a change in the composition of the two groups in society, and, because the poor group is getting smaller, to an increase in poor people's bias - $\frac{E_p}{E}$ will decrease. As described above, this means that even though people in the poor group have become poorer relative to*

[33] But it can also hold if $a \neq 0$, e.g. if $\Delta\underline{E}$ and $\Delta\bar{E}$ are such that $F\Delta\underline{E} + (1-F)(1-\beta(1-F))\Delta\bar{E} < 0$, then $dE_p$ is for sure $< 0$ because $\left[(1-F)^2\beta + f\beta(1-F)(\bar{E}-\hat{y})\right]d\hat{y}^* < 0$ if $\hat{y}^*$ decreases.

[34] (This depends on how much the median earner's income decreases relative to average income in the poor group.)

*the rich, because they misperceive average income more after the change in inequality, their perceived relative position might not have decreased, or might even have increased. Hence, whether support for redistribution increases or decreases in this case depends on the poor's degree of naivety and on how much the median income decreases due to the increase in inequality. In any case, even if the change in inequality is such that the demand for redistribution increases, the increase is smaller than would be expected in the framework of the Meltzer-Richard Model.*

- *If $a = 1$ and $\bar{E}$ increases ceteris paribus, then the equilibrium cutoff will go up. This leads to a change in the composition of the two groups in society, and, because the poor group is getting larger, to a decrease in poor people's bias - $\frac{E_p}{E}$ will increase. However, if the income distribution is sufficiently unequal such that the median earner is the decisive voter, the median voter's preferred tax rate will still be smaller than in the absence of segregation and misperceptions. However, the observed increase in support for redistribution might be larger if people are biased, because as $\bar{E}$ increases demand for redistribution increases for two reasons: the median voter is getting poorer relative to the average, and the median voter is becoming less biased and hence more aware of the prevailing inequality. While the first effect is larger if people are unbiased, the second effect is only present if people are biased, and together, the two effects might lead to a larger increase than in the absence of a bias.*

- *If $a = 1$ and both $\underline{E}$ decreases and $\bar{E}$ increases, the change in support for redistribution depends on whether the equilibrium cutoff increases or decreases. If $\Delta\bar{E} = -\Delta\underline{E}$, the equilibrium cutoff decreases if the income distribution is positively skewed. In this case the increase in support for redistribution will again be smaller than in the absence of misperceptions and we might even observe a decrease in support for redistribution.*

**Remark 5** *I do not have growth in my model, but my analysis would work in the same way if all variables would grow at a constant rate. In a model with growth, the case of $\bar{E}$ increasing and $\underline{E}$ decreasing would be translated into a situation where $\bar{E}$ increases a lot, while $\underline{E}$ stays constant (or increases only by a small rate), and we would see a decrease in the size of the poor group (corresponding to a decline in $\hat{y}^*$ with zero growth) if the distribution is sufficiently positively skewed. As Saez and Zucman (2016) point out, this constellation of high income growth of the rich accompanied by negligible growth rates of the bottom percentiles of the income distribution, is exactly what occurred during the past decades (at least in the US). Hence, my model can explain why, while inequality was increasing in the US over the past decades, people were, at least in the beginning, not demanding higher redistribution rates in response (if anything, then they were demanding lower redistribution rates, as documented by Kuziemko et al. (2015), who analyze the evolution of preferences for redistribution in the General Social Survey (GSS)).*

## A.9 Empirical Appendix

### A.9.1 Working with Amazon Mechanical Turk

For tax reasons, it is not possible for researchers living outside the United States to use Amazon Mechanical Turk directly. Therefore, I used the Amazon requester MTurkData to publish my survey via Amazon Mechanical Turk. They check the survey for compliance with Amazon's TOC, publish it on MTurk and deal with the payment of the workers afterwards.

The advantages and disadvantages of working with Amazon Mechanical Turk have been discussed by Kuziemko et al (2015) in their Webappendix. I agree with them that a major advantage of using MTurk is the speed of gathering responses: In my case, it took less than two hours to get 600 responses. There might in general be doubts about the quality of the responses, but it is possible to screen the MTurk workers based on their ratings for previous tasks. Using MTurk is also relatively cheap, as researchers design the survey themselves, instead of having it designed by a professional survey company. (Note also that I did not keep costs low at the expense of the respondents: they were all paid an hourly wage of 9 dollars.) One disadvantage of using MTurk is definitely that the obtained sample is usually not as representative as other, more expensive, online panel surveys (see below for a description of my own sample). However, as long as one keeps this in mind when interpreting the results, I think this is tolerable, especially when working with respondents from the United States, where MTurk is relatively well known and the pool of workers is therefore fairly representative.

### A.9.2 Sample characteristics

The sample is 83% White, 8.3% Black, 5.3% Asian and 1.5% Native American (the rest is "of other ethnicity"). Average age is 36.78, 44% of respondents are married. 68% are full- or part-time employees, 17% are self-employed and 13% are unemployed or not in the labour force. The respondents are very well educated, 63% have completed some kind of college degree. Hence, compared to other (more representative and commonly used) online panel surveys cited in the Webappendix of Kuziemko et al. my sample is younger, more educated and has fewer minorities. The household income distribution of the sample is roughly similar to the actual US household distribution (see Figures 7 and 8).

### A.9.3 Social Segregation: Description of Factor Analysis

In the survey, I ask several questions about people's colleagues at work, friends and family (spouse and siblings, if applicable). This is an attempt at identifying how diverse a person's social circle is. The underlying hypothesis is that an individual is more "socially segregated" the more homogenous and similar to herself her social circle is. However, it turned out that some of the questions were practically useless for my analysis in this relatively small sample: As less than half of the respondents are married, it turned out that using spouse characteristics

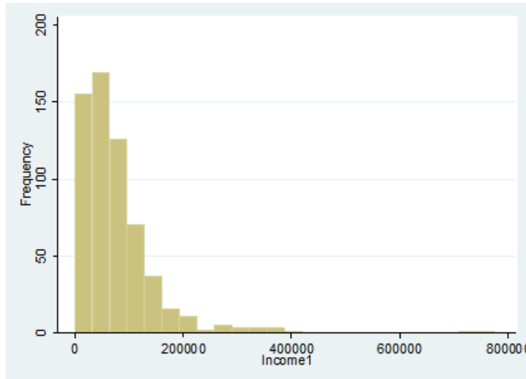


Figure 7: Sample household distribution

to categorize social segregation would exclude a big part of the sample, and a similar reason can be applied to sibling characteristics. I therefore decided to exclude those variables from my factor analysis. Furthermore, I excluded variables indicating whether friends or colleagues have the same mother tongue, because I figured out that these variables predominantly serve to identify Hispanics in the sample and do not provide much variation. Hence, the factor analysis utilizes four categorical variables classifying the similarity of friends' and colleagues' education and income level. The variables take on the value 0 if the respondent has answered that all of their friends/ colleagues are different to them in the respective area (e.g. the variable friends_educ is 0 if the respondent states that all of her friends have a different education level than herself) and is then increasing in the degree of similarity (i.e. 1 if most friends have different education levels,... up to 4 if all friends have the same education level as the respondent). Hence, the higher the value of each categorical variable, the higher the respondent's degree of social segregation.

The results of the factor analysis are presented in the main text.

## A.10 Case 2: Existence and uniqueness

In the following section I will analyze misperceptions of the form of Case 2 (see Section 5), where the poor people think average income in the rich group is higher than it actually is, while the rich people underestimate average income in the poor group.[35] Let me specifically assume that the belief function is such that:

$$\bar{E}_p(\hat{y}) = \beta(1 - F(\hat{y}))y_{\max} + (1 - \beta(1 - F(\hat{y}))\bar{E}(\hat{y}) \tag{30}$$

$$\underline{E}_r(\hat{y}) = \gamma F(\hat{y})0 + (1 - \gamma F(\hat{y}))\underline{E}(\hat{y}) \tag{31}$$

[35] Perhaps consumption of unrepresentative media could lead to such a bias: poor people watch "Celebrity Reality Shows" such as "Keeping up with the Kardashians" and conclude that rich people are very rich, while the rich read horror stories about deprivation in poor families and low standards of state schools.

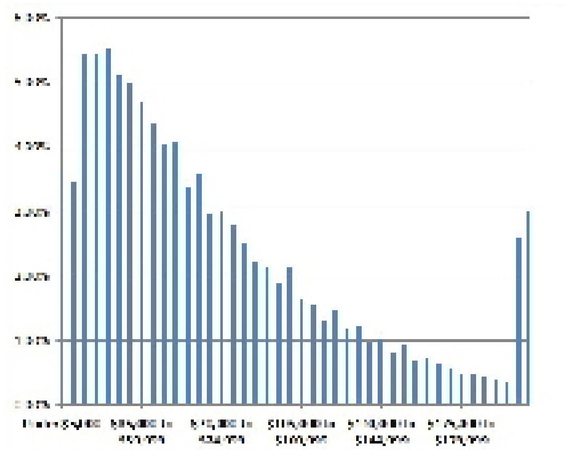

Figure 8: US household income distribution 2014 (Note: Figures for 2015 are not available yet, for a first approximation multiply the borders of the bins by average income growth to get an estimate for the 2015 income distribution).

| What is your current employment status | Freq. | Percent | Cum. |
|---|---|---|---|
| full-time employee | 345 | 57.21 | 57.21 |
| not in labor force (e.g. retired, ful.. | 49 | 8.13 | 65.34 |
| part-time employee | 66 | 10.95 | 76.29 |
| self-employed or small business owner | 104 | 17.25 | 93.53 |
| student | 10 | 1.66 | 95.19 |
| unemployed and looking for work | 29 | 4.81 | 100.00 |
| Total | 603 | 100.00 | |

Figure 9: Types of employments held by sample respondents

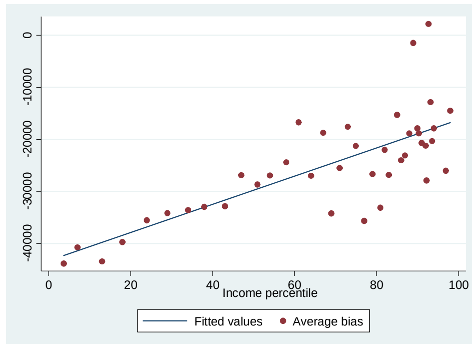


Figure 10: Average misperception of the mean household income by income percentiles (Bias = correct average income - perceived average income)

Analogous to Section 6, $\beta$ and $\gamma$ parameterize the "naivity" of the poor and the rich respectively, and if $\beta$ resp. $\gamma$ is 0 agents have no misperceptions. As for Case 1, the functional form of $\bar{E}_p$ and $\underline{E}_r$ implies that the misperceptions are more severe, the smaller the part of the distribution they can fully observe. It is straightforward to see that $\bar{E}_p \geq \bar{E}$ and $\underline{E}_r \leq \underline{E}$. The misperceptions converge to the truth monotonically, and hence if $U$ is linear in both arguments (as I will henceforth assume), the equilibrium condition becomes

$$\hat{y}^* \left[\bar{E}(\hat{y}^*) - \underline{E}_r(\hat{y}^*)\right] = \hat{y}^* [\bar{E}_p(\hat{y}^*) - \underline{E}(\hat{y}^*)]$$

and the unique interior equilibrium cutoff satisfies

$$\bar{E}(\hat{y}^*) - \underline{E}_r(\hat{y}^*) = \bar{E}_p(\hat{y}^*) - \underline{E}(\hat{y}^*)$$

Plugging in the functional form of the misperceptions, (30) and (31), yields

$$\beta(1 - F(\hat{y}^*))(y_{\max} - \bar{E}(\hat{y}^*)) = \gamma F(\hat{y}^*)\underline{E}(\hat{y}^*) \tag{32}$$

This equation indirectly characterizes $\hat{y}^*$.[36]

[36] Existence is confirmed by seeing that as $\hat{y} \to 0$ the LHS goes to $\beta(y_{\max} - E)$ whereas the LHS goes to 0, while at $\hat{y} \to y_{\max}$ the LHS (0) is smaller than the RHS ($\gamma E$). As the expressions on both sides are continuous functions of $\hat{y}$, there must be a cutoff $\hat{y}^* \in (0, y_{\max})$ such that both sides are equal. Rewriting equation (32) using $a = \frac{\beta}{\gamma}$ I get

$$F\underline{E} + a(1 - F)\bar{E} = a(1 - F)y_{\max} \tag{33}$$

To confirm that there can only be one $\hat{y}^*$ satisfying this equation, I employ a single-crossing argument: Determine the slope of the LHS and the RHS by taking the derivative with respect to $\hat{y}$ on both sides. This yields

$$f\hat{y} - af\hat{y}$$

for the LHS and

$$-afy_{\max}$$

for the RHS. Clearly, for any $\hat{y}$ we have

$$f(\hat{y} - a\hat{y}) > -afy_{\max}$$

If poor and rich people are equally naive, then $\beta = \gamma$ and equation (32) simplifies to

$$1 - F(\hat{y}^*) = \frac{E}{y_{\max}}$$

In this case, it is immediate to see that the equilibrium is unique, as the RHS does not vary with $\hat{y}$ and $F$ is strictly increasing and hence there will be only one $\hat{y}^*$ that satisfies this equation. Moreover, the cutoff is decreasing in $\frac{E}{y_{\max}}$, i.e. it is higher the larger the difference is between maximum and average income. Furthermore, the equilibrium cutoff must lie above the median, because the income distribution is positively skewed and hence $\frac{E}{y_{\max}}$ will be smaller than $\frac{1}{2}$ and therefore $F(\hat{y}^*)$ must be larger than $\frac{1}{2}$. As in Case 1, also here the equilibrium cutoff only depends on $a$, not on $\beta$ and $\gamma$ individually.

### A.10.1 Application for Case 2: Housing and education

Take a segregated city in the US and suppose the biases of Case 2 are at work there. This would imply that in rich neighborhoods, the average income of the poor (and hence the average benefit of mixing with them) is underestimated and hence the rich are willing to pay more to segregate from the poor than their actual benefit from sorting. (Equally, the poor are also willing to pay more to mix with the rich than in the unbiased model). This implies that for example housing prices in rich neighborhoods (if this is what we interpret the sorting fee $b$ to be) would be exaggeratedly high, or that fees for private schools are very high.[37] How exaggerated these prices are depends on the degree of naivety of the poor versus the rich and the shape of the income distribution, as these two factors determine the cutoff $\hat{y}^*$ and hence the sorting fee $b$ and the severity of the misperceptions. The sorting fee $b$ is given by

$$b = \hat{y}^* \left[\bar{E} - \underline{E}_r\right]$$

or equivalently

$$b = \hat{y}^*[\bar{E}_p - \underline{E}]$$

and hence is increasing in the equilibrium cutoff and in the perceived benefits of sorting, $\bar{E} - \underline{E}_r$ resp. $\bar{E}_p - \underline{E}$. As

$$\bar{E} - \underline{E}_r = \bar{E} - (1 - \gamma F)\underline{E}$$

(because $f$ is a pdf and therefore always positive). At any point $\hat{y}$ the slope of the LHS is larger than the slope of the RHS. This holds both when $a < 1$ (in which case the LHS is increasing in $y$, while the RHS is decreasing) and when $a > 1$ (in which case both sides are decreasing, but the RHS slope is steeper). Hence, the same must hold at any point where the two sides cross. This implies that the RHS must always cut the LHS from above, which means that the two can only cross once.

[37] Note that I have deliberately chosen the example of a US segregated city, because in the US, housing and schooling are closely connected because children have to attend local schools, and so basically house prices also reflect the quality of local schools, given that parents want their kids to attend the best schools possible.

I find that $b$ is increasing in $\bar{E}$, decreasing in $\underline{E}$ and increasing in the degree of naivety, $\gamma$.

Similar to my analysis for Case 1, I could also examine how a change in inequality will affect the equilibrium cutoff $\hat{y}^*$ and the sorting fee $b$ (which would in this case correspond to house prices). I defer this analysis to a later paper of mine. In the following sections, I will examine how Case 1- and Case 2-type misperceptions differ in terms of their implications for welfare and monopolist revenue (if the sorting technology is offered by a profit-maximizing monopolist).

## A.11 Welfare comparison Case 1, Case 2 and unbiased

If society is segregated with cutoff $\hat{y}$, total welfare can be calculated as[38]

$$W_S = \int_0^{\hat{y}} y\underline{E}f(y)dy + \int_{\hat{y}}^{y_{\max}} y\bar{E}f(y)dy - (1 - F(\hat{y}))b \tag{34}$$

If people are unbiased, the sorting fee $b$ must satisfy

$$b = \hat{y}(\bar{E} - \underline{E})$$

If people are biased according to Case 1, where both groups underestimate the benefits of sorting, the sorting fee at the equilibrium cutoff is

$$b = \hat{y}^*(\bar{E} - \underline{E}_r) = \hat{y}^*(\bar{E}_p - \underline{E})$$

As

$$\hat{y}^*(\bar{E} - \underline{E}_r) = \hat{y}^*(\bar{E} - \underline{E}) - \hat{y}^*\gamma F(\hat{y}^* - \underline{E}) < \hat{y}^*(\bar{E} - \underline{E})$$

$b$ in Case 1 is smaller than the sorting fee in the unbiased case for the same cutoff $\hat{y}^*$. Hence, welfare under sorting with misperceptions according to Case 1 delivers a higher total welfare than unbiased sorting at the same cutoff.

If people are biased according to Case 2, where both groups overestimate the benefits of sorting, the sorting fee at the equilibrium cutoff is again

$$b = \hat{y}^*(\bar{E} - \underline{E}_r) = \hat{y}^*(\bar{E}_p - \underline{E})$$

However, in Case 2 we get

$$\hat{y}^*(\bar{E} - \underline{E}_r) = \hat{y}^*(\bar{E} - \underline{E}) + \hat{y}^*\gamma F\underline{E} > \hat{y}^*(\bar{E} - \underline{E})$$

and hence the sorting fee is higher than in the unbiased case for the same cutoff $\hat{y}^*$.

[38] As in Levy and Razin (2015), total welfare from a particular partition takes into consideration the sorting fee paid (as deadweight loss to society, or benefitting only a negligible proportion of society). If the sorting fee would not be considered, perfect sorting would always be efficient, because the utility from a match is supermodular (see Becker (1974)).

**Proposition 25** *If people are biased according to Case 1, where both groups underestimate the benefits of sorting, total welfare of sorting at the equilibrium cutoff $\hat{y}^*$ is higher than unbiased sorting at the same cutoff.*

**Proposition 26** *If people are biased according to Case 2, where both groups overestimate the benefits of sorting, total welfare of sorting at the equilibrium cutoff $\hat{y}^*$ is lower than unbiased sorting at the same cutoff.*

### A.11.1 Welfare and increasing inequality

Suppose that the sorting technology is offered by a benevolent social planner who wants to maximize welfare. When deciding whether or not to offer the sorting technology, she will evaluate total welfare under no sorting and compare it to total welfare with two groups for the equilibrium cutoff $\hat{y}^*$.

If there is no segregation in society, no sorting fees are paid and everybody interacts with everybody else. Hence a person with income $y_i$ gets utility $y_i E$ and total welfare in society is

$$W_{NS} = \int_0^{y_{\max}} yE(y)f(y)dy = E^2$$

Welfare of sorting at some cutoff $\hat{y}$ is given by (34). If people are unbiased, the difference between welfare of sorting at some $\hat{y}$ and welfare of no sorting can be written as

$$W_S - W_{NS} = F\underline{E}^2 + (1-F)\bar{E}^2 - \hat{y}^*(1-F)(\bar{E} - \underline{E}) - E^2 \tag{35}$$

Levy and Razin (2015) show that expression (35) can be written as

$$(1-F)(\bar{E} - \underline{E})(\bar{E} - \hat{y}^* - E)$$

which will be positive for all $\hat{y}^*$ iff

$$\bar{E} - E > \hat{y}^* \quad \forall \hat{y}^*$$

This holds if $F(y)$ is *new worse than under expectations (NWUE).*

If people are biased according to Case 1, welfare of sorting at $\hat{y}^*$ can be rewritten as

$$W_S^1 = F\underline{E}^2 + (1-F)\bar{E}^2 - \hat{y}^*(1-F)(\bar{E} - \underline{E}) + \hat{y}^*(1-F)\gamma F(\hat{y}^* - E)$$

Hence, in this case the welfare difference between a situation with sorting (at equilibrium cutoff $\hat{y}^*$) and a situation with no sorting is

$$W_S^1 - W_{NS} = F\underline{E}^2 + (1-F)\bar{E}^2 - \hat{y}^*(1-F)(\bar{E} - \underline{E}) + \hat{y}^*(1-F)\gamma F(\hat{y}^* - \underline{E}) - E^2 \tag{36}$$

Compared to the unbiased case, the welfare difference now contains the extra term $\hat{y}^*(1-F)\gamma F(\hat{y}^* - \underline{E})$, which is positive. Hence, $F(.)$ being NWUE is a sufficient condition for welfare being higher under sorting than under no sorting (for any cutoff) if people are biased according to Case 1.

**Corollary 7** *If people are biased according to Case 1 and F is NWUE, a benevolent (utilitarian) social planner prefers sorting (at any cutoff) to no sorting.*

For the particular case where the poor and the rich are equally naive, expression (36) can be further simplified using the fact that $\hat{y}^* = E$ if $a = 1$.

$$\begin{aligned} W_S^1 - W_{NS} &= F\underline{E}^2 + (1-F)\bar{E}^2 - E(1-F)(\bar{E} - \underline{E}) + E(1-F)\gamma F(E - \underline{E}) - E^2 \\ &= (1-F)(\bar{E} - \underline{E})(\bar{E} - 2E) + E(1-F)\gamma F(E - \underline{E}) \\ &= (E - \underline{E})\left(\bar{E} - 2E + \gamma E(1-F)F\right) \end{aligned}$$

Hence

$$\begin{aligned} W_S^1 - W_{NS} > 0 &\iff (E - \underline{E})\left(\bar{E} - 2E + \gamma E(1-F)F\right) > 0 \\ &\iff \bar{E} > E(2 - \gamma(1-F)F) \qquad (37) \end{aligned}$$

Now suppose that (37) is not satisfied at first, but then inequality increases in the sense of a mean preserving spread that keeps $F(E)$ and $E$ constant, increases $\bar{E}$ and decreases $\underline{E}$. It is straightforward to see that this increases the RHS of (37), while leaving the LHS constant. Thus, offering segregation can become efficient if inequality increases.

**Proposition 27** *If people are biased according to Case 1, an increase in inequality in the sense of a mean preserving spread that keeps $F(E)$ constant, increases $\bar{E}$ and decreases $\underline{E}$, increases the welfare difference between a situation with segregation and a situation without segregation. Hence, such an increase in inequality can make it desirable for a benevolent planner to switch from a society without segregation to a society with segregation.*

Comparing the situation where inequality increases in Case 1 to the situation of increasing inequality when people are unbiased at the same cutoff ($E$, which is the equilibrium cutoff in Case 1 if $a = 1$), I find that sorting at $E$ will be efficient in the unbiased case iff the income distribution is such that

$$\bar{E} > 2E.$$

In Case 1, sorting is efficient already at a lower degree of inequality (measured as $\bar{E} - E$, or equivalently $\bar{E} - \underline{E}$), namely if the income distribution is such that

$$\bar{E} > E(2 - \gamma(1-F)F).$$

The reason is that for the same cutoff welfare is always higher in Case 1 than if people are unbiased, because the sorting fee is lower, hence there will be degrees of inequality where sorting is efficient in Case 1 but not efficient if people are unbiased.

If people are biased according to Case 2, welfare can be written as

$$W_S^2 = F\underline{E}^2 + (1-F)\bar{E}^2 - (1-F)\hat{y}^*(\bar{E} - \underline{E}) - (1-F)\hat{y}^*\gamma F\underline{E}$$

and it is immediate to see that for any cutoff, welfare in Case 2 is lower than in Case 1 and in the unbiased case. F being NWUE is a necessary and sufficient condition for sorting to be efficient (at any cutoff) in the unbiased case (and a sufficient condition in Case 1), but in Case 2 NWUE is not enough for sorting at any cutoff to yield higher welfare than no sorting, because the sorting fee is higher than in the unbiased case.

The difference between welfare of sorting at $\hat{y}^*$ in Case 2 and welfare of no sorting can be written as

$$W_S^2 - W_{NS} = (1-F)(\bar{E} - \underline{E})(\bar{E} - \hat{y}^* - E) - (1-F)\hat{y}^*\gamma F\underline{E}$$

We can now again look what happens to this difference after a mean-preserving spread and compare Case 2 where the equilibrium cutoff is at $E$ to Case 1 and unbiased sorting at $E$ : If the income distribution is such that $E$ is the equilibrium cutoff in Case 2, the welfare benefit from sorting compared to no sorting is

$$(1-F)(\bar{E} - \underline{E})(\bar{E} - 2E) - (1-F)E\gamma F\underline{E}.$$

If there is a mean-preserving spread that leaves $F(E)$ constant, increases $\bar{E}$ and decreases $\underline{E}$, the first summand will increase, while the term that is substracted will decrease, and therefore the welfare benefit from sorting will increase, and can go from positive to negative. However, compared to the unbiased case and Case 1, this will happen only for larger degrees of inequality (as measured by $\bar{E} - \underline{E}$), because the sorting fee is higher.

**Proposition 28** *If $a = 1$ and the income distribution is such that the equilibrium cutoff is $\hat{y}^* = E$ in Case 2, an increase in inequality in the form of a mean-preserving spread that leaves $F(E)$ constant, increases $\bar{E}$ and decreases $\underline{E}$ makes sorting at $E$ efficient (compared to no sorting) in Case 1 already for lower levels of inequality (as measured by $\bar{E} - \underline{E}$) than in the unbiased case and it makes sorting at $E$ efficient in the unbiased case already for lower levels than in Case 2. In other words, there exist levels of inequality such that sorting at $E$ is efficient in Case 1 but not in the other cases. There exist levels of inequality such that sorting at $E$ is efficient in Case 1 and in the unbiased case, but not in Case 2.*

## A.12 Monopolist profit comparison of the three cases

Suppose a profit-maximizing monopolist who has a fixed cost $c > 0$ of offering the sorting technology can decide whether or not to become active. If people are biased according to Case 1, the monopolist's profit from offering sorting is

$$\hat{y}^*(\bar{E}(\hat{y}^*) - \underline{E}_r(\hat{y}^*))(1 - F(\hat{y}^*)) - c$$

Given that the equilibrium cutoff is at $E$ and substituting for $\underline{E}_r$, this can be rewritten as

$$E(E - \underline{E}(E))[1 - \gamma F(E)(1 - F(E))] - c \tag{38}$$

Suppose that initially the income distribution is such that

$$E(E - \underline{E}(E))[1 - \gamma F(E)(1 - F(E))] - c < 0$$

and hence the monopolist would prefer to stay out of the market. If inequality increases (again in the sense of a mean-preserving spread of the income distribution, which decreases $\underline{E}$ and increases $\bar{E}$, while leaving $E$ and $F(E)$ constant), $E - \underline{E}$ increases. This means that if the increase in inequality is sufficiently large, the profits from offering the sorting technology will become positive and the society will become segregated. Thus, a large enough increase in inequality will lead to economic segregation.

If people are biased according to Case 2, the monopolist's profit from offering the sorting technology is again

$$\hat{y}^*(\bar{E} - \underline{E}_r)(1 - F(\hat{y}^*)) - c$$

but now people overestimate the benefits of sorting and hence this expression can be rewritten as

$$\hat{y}^*(\bar{E}(\hat{y}^*) - (1 - \gamma F(\hat{y}^*))\underline{E}(\hat{y}^*))(1 - F(\hat{y}^*)) - c$$

It depends on the shape of the income distribution whether the monopolist's profit increases or decreases due to a mean-preserving spread. Remember that the equilibrium cutoff in Case 2 if $a = 1$ is given by

$$1 - F(\hat{y}^*) = \frac{E}{y_{\max}}$$

Hence, the equilibrium cutoff need not be at average income $E$ in this case, the exact location of $\hat{y}^*$ depends on the income distribution. This means that in general, a mean-preserving spread will not change only $\bar{E}$ and $\underline{E}$ but also the equilibrium cutoff. Therefore, the overall effect of a mean-preserving spread on the monopolist's profits are ambiguous: $\bar{E} - \underline{E}_r$ increases, but the equilibrium cutoff may go up or down and what happens to the overall sorting fee $b$ and to the monopolist's profits depends on the income distribution. Thus, we cannot in general compare whether the monopolist will be quicker to enter than in Case 1 if inequality increases. However, if the income distribution is such that the equilibrium cutoff is also at $E$ in Case 2, then a mean-preserving spread (that leaves $F(E)$ constant) will not affect the location of the equilibrium cutoff and the monopolist's profits will increase due to a mean-preserving spread. Moreover, the monopolist will offer the sorting technology for lower degrees of inequality (as measured by $\bar{E} - \underline{E}$) than the monopolist in Case 1, because her revenue $\hat{y}(1 - F(y))b$ is higher for any cutoff (and therefore also for $\hat{y} = E$) than in Case 1, because the sorting fee $b$ is higher.

If people are unbiased, the sorting fee for a given cutoff $\hat{y}$ amounts to

$$b = \hat{y}(\bar{E}(\hat{y}) - \underline{E}(\hat{y}))$$

and hence the monopolist's profits from offering the sorting technology at cutoff $\hat{y}$ are

$$\Pi(\hat{y}) = (1 - F(\hat{y}))\hat{y}(\bar{E}(\hat{y}) - \underline{E}(\hat{y})) - c$$

It is straightforward to see that for any cutoff $\hat{y}$ the monopolists profits will be lower than in Case 2 and higher than in Case 1.

If people are unbiased, the monopolist can set the cutoff anywhere in $Y$ and will therefore set it such that her profits are maximized. This means that at the optimal cutoff we need[39]

$$\Pi'(\hat{y}) = 0$$

which can be rewritten as

$$\bar{E}(\hat{y}) - E = \hat{y}\frac{f}{1 - F(\hat{y})}(\hat{y} - \underline{E}(\hat{y})) \tag{39}$$

For reasons of comparison, let me assume that the income distribution is such that the optimal cutoff is at average income $E$. Plugging $\hat{y} = E$ into 39 and rearranging, I find that for average income to be the optimal cutoff, the income distribution must be such that

$$\frac{F(E)}{f(E)} = E$$

Suppose that income is initially distributed relatively equally, such that the difference between average income of the rich and average income of the poor is small, i.e. $\bar{E} - \underline{E}$ is low, and the monopolist's profits are negative. If income inequality increases in the form of a mean-preserving spread that leaves $F(E)$ and $f(E)$ constant then the equilibrium cutoff will not change, but $\bar{E} - \underline{E}$ will increase, and hence also the monopolist's profits. If the mean-preserving spread is large enough, the monopolist will find it profitable to offer the sorting technology. Offering sorting will become profitable for smaller degrees of inequality than in Case 1 and for larger degrees of inequality than in Case 2.

This analysis yields the following Propositions:

**Proposition 29** *Let people be biased according to Case 1 and suppose that the income distribution is initially such that a profit maximizing monopolist with fixed costs $c > 0$ does not find it profitable to offer the sorting technology. Then for any $c > 0$ there exists a mean-preserving spread of the income distribution such that the monopolist's profits become positive.*

**Proposition 30** *Let people be biased according to Case 2 and let the income distribution be such that the equilibrium cutoff is at average income. Suppose that the income distribution is initially such that a profit maximizing monopolist with fixed costs $c > 0$ does not find it profitable to offer the sorting technology. Then for any $c > 0$ there exists a mean-preserving spread of the income distribution such that the monopolist's profits become positive.*

[39] It is straightforward to see that this maximization problem has an interior solution, because $\Pi(0) = \Pi(y_{\max}) = -c$ whereas any interior $\hat{y}$ yields $\Pi(\hat{y}) > -c$.

**Proposition 31** *Let people be unbiased and let the income distribution be such that the monopolist s optimal profit is at average income. Suppose that the income distribution is initially such that a profit maximizing monopolist with fixed costs* $c > 0$ *does not find it profitable to offer the sorting technology. Then for any* $c > 0$ *there exists a mean-preserving spread of the income distribution such that the monopolist's profits become positive.*

**Proposition 32** *The monopolist's profits will be higher in Case 2 than in Case 1 for any cutoff* $\hat{y}^*$ *and for any degree of inequality of the income distribution. Therefore, as income inequality increases, the monopolist's profits in Case 2 will become positive already for smaller degrees of inequality than necessary for her profits in Case 1 to be positive.*

**Proposition 33** *For any income distribution, the profits for the unbiased case lie in between the profits for Case 2 and Case 1 (for the same cutoff). Therefore, as income inequality increases, the monopolist's profits in the unbiased case will become positive already for smaller degrees of inequality than necessary for her profits in Case 1 to be positive. However, higher degrees of inequality are needed for her profits to be positive than in Case 2.*