Essays in the Political Economy of Redistribution and Nation Formation

Sabine Flamand

Supervised by Caterina Calsamiglia

Submitted to
Departament d'Economia i d'Història Econòmica
at
Universitat Autònoma de Barcelona
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

May 2012

Acknowledgments

I am mostly indebted to Caterina Calsamiglia, who always believed in me from the very beginning, and who pushed me in the right direction at the very right moment. As I already told her once, that day I came to her office to ask her to be my advisor was one of the best decisions I ever took.

I would like to thank her for her continuous support, her great availability, and her priceless feedback. I am also grateful to her for her infinite patience, and for always putting me back on track whenever I was doubting. She allowed me to think and evolve as a researcher, and helped me push my boundaries away. I learnt a lot from her about research and economic thinking, but about so much more than that as well, and I hope I will continue learning from her for many years in the future. She has been a great advisor, but also an amazing coach, a friend I can trust, and sometimes even a therapist.

I simply want to thank Caterina, in fact, for being Caterina.

I would like to thank Clara Ponsatí, Andrea Mattozzi, Susana Peralta, Enriqueta Aragonès, and Humberto Llavador for kindly accepting to be members of the committee. In particular, I am grateful to Clara, Andrea and Enriqueta for their constructive feedback, and to Susana for trusting me, and for her time and involvement to help me make my projects materialize. I also want to thank Joan María Esteban, Carmen Beviá, Hannes Mueller, Pedro Rey, as well as the other IDEA faculty members for the inspiring discussions we had, and for their constructive suggestions. Thanks also to Carmen for her flexibility and understanding as the director of the program when I was at a crucial crossroads, to Angels López and Mercè Vicente for their help during these years, and to Laura Mayoral for her flexibility in the very last stretch.

I am greatly indebted to my classmates, who are not only great colleagues, but also amazing friends. Thanks to Philippos, who I got to really know way too late, for our great partnership in all dimensions during this last year, and for being the best travelling companion ever. Thanks to Orestis, from whom I learnt much more than he thinks, for his sense of humor, and without whom this experience would have been very different. Thanks to Riste, for transmitting me part of his great life philosophy. Thanks also to Javi, who left Barcelona too early, for his serenity. Thanks to them, and to all my IDEA colleagues, who made the PhD experience truly entertaining sometimes. I also want to thank my classmates in Brussels, Charles Plaigin, Julien Ravet and Laurent Bouton, with whom I really got to learn and understand Economics.

I am forever indebted to Anne, *mi compañera de vida*, for being my sister, my mother, my best friend, and much more than that during all these years. I am convinced deep inside that no matter the distance, we will always be part of each other's life. I want to thank Magali and Anaïs, for making my life in Barcelona so colorful and a bit more like holidays, and for standing by my side during the (few) worst and (many) best moments. Thanks to Matthew, for our endless philosophical conversations accompanied with countless beers, and for being one of the greatest persons I ever met. Thanks to Olivier, who is probably the only person on earth whining more than I do, for his frankness in any possible situation. Thanks to Patricia, for being such a happy mess, and for always being there for me. Thanks to Aurélia, for this amazing party we had for our 30's, and from which I am still trying to recover. Thanks to Thomas, for always being himself in such a light way. Thanks to them, and to Aurélie, Stéphane, Stan, Régis, Vanessa, Eva, and many others, for making these years in Barcelona one of the best parts of my life so far.

I also want to thank my friends in Brussels, Renaud, Manu, Maude, Virginie, Sigiri, Diana, Céline, Carole, Caro and many others (although they always make fun of me being still a student above the age of 30) for their irreplaceable friendship since all those years, and who I miss deeply since I left. The moments we share all together when I come home are pure happiness, and are priceless.

A special thank should go to Virginie, whose entertaining weekly *topos* kept me informed about every single gossip related to Brussels.

I am greatly indebted to my parents, Jacques and Françoise, for their unconditional support, and for always doing whatever they could to make my wishes become reality. I am greatful to my brother, Jean-François, for his wisdom. This thesis is dedicated to them, to my sisters, Cécile, Stéphanie and Lucie, and to Maya and Antoine, the two newest arrivals, who are growing up way too fast. Thanks also to my grand-parents, Georges and Lucette, who are really special to me, and to my uncle Georges-André, my best *apéro* partner, for always finding the words to make me laugh.

Finally, thanks to Marta, an eternal optimist, for her love, and for making me happy during this last and intense year.

Contents

Chapter 1

General Introduction

This dissertation mainly explores the effects of income inequality and other forms of heterogeneity —and in particular at the regional level— on the individual willingness to implement general redistribution schemes, and to form part of a political union.

We start by analyzing the effect of different types of altruism on individual preferences for redistribution, and investigate what kind of political equilibrium emerges from the prevalence of different altruistic motives among the population for a given distribution of income.

We then use such a model of voting on redistribution with altruism in order to study the choice between a centralized versus decentralized system of redistribution, assuming two-sided heterogeneity between regions, namely average income and group identity.

Finally, we study the relevance of using partial decentralization of both public expenditures and revenues as a separatist conflict-mitigating strategy, with a particular focus on the role of interregional income inequality in answering this question.

In the second chapter, we introduce heterogeneous social preferences in a standard model of voting on a redistributive parameter in a direct democracy. In particular, and in accordance with experimental evidence, we assume that selfish, rawlsian and utilitarian voters coexist with given proportions. We characterize implicitly the unique political equilibrium of this economy, and prove its existence for any positively skewed income distribution. It turns out that the level of redistribution in the heterogeneous economy may be either lower or higher than in the selfish one.

Furthermore, we show that slight variations in the relative proportion of a given type may lead to very important changes in the extent of redistribution, and we illustrate the implications this may have in the context of the political economy of border formation. In particular, we show that, even when there are no income differences whatsoever between regions, the fact that the relative proportion and/or distribution of a given type differs across regions is very likely to prevent voluntary regional integration.

Finally, we investigate the theoretical implications of the model regarding the link between inequality and redistribution, and show that it yields different predictions than the standard model with self-interested voters. In particular, an increase in poverty is very likely to increase redistribution. Furthermore, this is also true for a mean-preserving spread leaving the median income unaffected, although it has no effect whatsoever on redistribution in the traditional selfish economy.

In the third chapter, we study the choice between centralized and decentralized redistribution in a political economy model, assuming regional heterogeneity regarding both average income and group identity. While centralizing redistribution allows for a potentially beneficial pooling of national resources, it may also decrease the degree of solidarity in the society as a result of group loyalty.

In this context, we show that total welfare maximization is closely linked to the minimization of inequality both within and between regions. In particular, we show that total redistribution is best, that is, such that all individuals end up with the same final consumption, both within and between regions. It turns out, then, that unlike in the traditional fiscal federalism literature, the first-best solution is only (theoretically) attainable under centralized redistribution whenever one region is richer than the other (i.e. whenever regions are not identical).

Analyzing separately two particular cases under direct democracy —no interregional inequality and no group loyalty— allows us to highlight the existence of a scale effect and a pooling effect of centralized redistribution, respectively. The scale effect relates to the fact that, whenever group loyalty is not perfect, and independently of the existence of interregional transfers, individuals are willing to redistribute more (i.e. pay more taxes) when redistribution is centralized, just because the latter is implemented at a bigger scale. In contrast, what we call the pooling effect is the fact that in the absence of group loyalty, all individuals strictly prefer centralization to decentralization, since the former allows to pool national resources so as to homogenize consumption of poor and rich individuals throughout the country (i.e. centralization allows to redistribute both within and between regions). In both cases, centralization pareto-dominates decentralization, from which it follows that the rationale for decentralization only arises when both sources of regional heterogeneity are present. This means, in turn, that Oates' (1962) decentralization theorem does not hold in our political economy approach of redistribution with group loyalty.

Finally, allowing for voluntary interregional transfers under decentralization, we show that, due to the presence of free-riding, centralization always welfare-dominates decentralization with transfers whenever such transfers occur. Furthermore, it is not even generally true that allowing for such transfers is welfare-increasing under decentralization.

Finally, in the fourth and last chapter, we study the use of partial decentralization of both public expenditures and revenues as a way to avoid wasteful secessionist conflict in the presence of income disparities between regions, and thus implicit interregional transfers under unification.

Although decentralization allows regional governments to better target local preferences, it also implies some losses in terms of economies of scale in the provision of public goods. Furthermore, decentralization of public revenues (i.e. fiscal autonomy) also has a tendency to exacerbate interregional inequality, which may have both stabilizing and destabilizing effects.

In order to study the relevance of partial decentralization as a conflict-mitigating strategy in this context, we use a standard political economy model of nation formation, which we extend for the presence of interregional inequality, and an explicit (separatist) conflict technology. Finally, we introduce the possibility of partially decentralizing the union as an alternative to such conflict.

We show that, even though an increase in inequality fuels conflict in both regions, the probability of a secession occurring through conflict might be either increasing or decreasing in inequality, depending on whether unifying is socially efficient or not. It follows that, on the one hand, the range of decentralization levels such that the peaceful —decentralized— outcome is politically sustainable always increases with inequality, while on the other hand, the particular level of decentralization that is implemented in the shadow of conflict may be either decreasing or increasing in inequality. Finally, we show that when decentralization is an institutionally irreversible process, it cannot prevent secessionist conflict for any level of inequality.

Chapter 2

Heterogeneous Social Preferences in a Model of Voting on Redistribution

2.1 Introduction

Recent experimental evidence indicates that individual concerns for fairness and altruism can explain a range of economic phenomena which are not in accordance with the traditional selfishness assumption. Other-regarding preferences may take several forms: altruism, inequity aversion, and reciprocity (Fehr and Schmidt (2006)). Furthermore, there is strong evidence that individual concerns for fairness are not homogeneous. Rather, some individuals may be more altruistic than others, some are simply totally selfish, but it may also well be the case that people derive utility from different types of altruism.

In this paper, we investigate this latter possibility by introducing heterogeneous social preferences in a standard model of voting on a redistributive parameter. More specifically, we assume that three different types of voters coexist with given proportions: selfish, rawlsian, and utilitarian. Tyran and Sausgruber (2006) and Ackert et al. (2007a and b) have shown in their experiments that individuals demonstrate concerns for others when voting for a redistributive policy. Furthermore, there is also strong evidence that a significant proportion of individuals exhibit either selfish, rawlsian or utilitarian preferences.

According to the experiments conducted by Andreoni and Miller (2002), people differ on whether they care about fairness at all, and when they do, the notion of fairness they employ differs widely, ranging from rawlsian to utilitarian. They report that 44% of their subjects are completely selfish, 35% exhibit egalitarian preferences, and 21% of the subjects can be classified as surplus maximizers. Fisman et al. (2007) and Iriberri and Rey-Biel (2008) also find similar proportions[1]. Bolton and Ockenfels (2002), investigating the trade-off between efficiency and equity in a voting experiment,

[1]Note that there is also some evidence that people may behave jealously, and that a minority of people is competitive. However, Erlei (2008), allowing for this latter type of preferences in his analysis of 43 games, finds that it does not improve the predictive success of the approach.

find that, as a social good, equity is in greater demand than is efficiency. According to their results, about twice as many people deviate from pure self-interest for equity than for efficiency. Finally, Erlei (2008, p. 436), using a simple model which combines the basic ideas of different models of social preferences, and especially allowing for heterogeneity, applies it to 43 games and concludes that '*[...] models of social preferences are particularly powerful in explaining behavior if they are embedded in a setting of heterogeneous actors with heterogeneous (social) preferences*". That is, he finds that the analytical combination of social preferences, together with heterogeneity in these preferences, greatly furthers the understanding of real behavior in laboratory experiments.

The standard approach on redistribution through the voting process comes from the models of Romer (1975), Roberts (1977) and Meltzer and Richards (1981). The commonly used name for this class of models is the RRMR model. It consists in a general equilibrium model which assumes purely selfish individuals who differ with respect to their productivity, which in turn determines their income. It predicts that the extent of redistribution is determined by the median-income voter, with higher-income individuals supporting strictly less redistribution than lower-income ones. More specifically, the median voter, being selfish, supports a positive level of redistribution only to the extent that he's poorer than average —and thus benefits himself from redistribution. Consequently, the model predicts that higher inequality, as measured by the mean-to-median income ratio, translates into more redistribution in equilibrium.

In this paper, we examine the theoretical implications of assuming heterogeneous social preferences in the latter traditional framework of voting on redistribution. In particular, we assume that selfish, rawlsian and utilitarian voters coexist with given proportions, and given weights associated to altruistic motives relatively to self-interest in their utility function. Given the significant experimental evidence highlighting the coexistence of those three distinct types of individuals, we believe it is important to investigate what kind of political equilibrium would arise from the interactions between them (if any), and what are its properties. That is, we aim at determining what are the equilibrium redistribution outcomes resulting from the prevalence of different altruistic motives among the voters. Observe that we abstract from the possibility of reciprocity, as we believe that such behavior is more likely to occur in strategic settings, where players can directly affect each other's payoffs (in the case of bilateral interactions, for example), than in a voting context with a large electorate.

Our findings are as follows: first, we characterize implicitly the (strictly positive) Condorcet winner tax rate of the heterogeneous economy, and prove its existence for any positively skewed income distribution. Second, we show that, while an increase in the relative proportion of rawlsian (utilitarian) voters always yields more (less) redistribution, an increase in the relative proportion of selfish voters has an ambiguous effect on the equilibrium tax rate. Indeed, given that the redistributive preferences of the selfish voters have an intermediate position between the rawlsian and utilitarian ones for any level of income, it turns out that the level of redistribution in the heterogeneous economy may be either lower or higher than in the selfish one. Third, using simulations of the model, we show that varying slightly the relative proportion of a given type can have a very large impact on the redistributive outcome when voters attach a high relative weight to altruistic motives in their utility function. In turn, we illustrate with some examples the implications of this latter fact in the context of (the political economy of) border formation. In particular, we show that, even though there are no income differences whatsoever between regions, the fact that the relative proportion and/or distribution of a given type differs across regions is very likely to prevent voluntary regional integration. Finally, we show that the standard prediction regarding the link between inequality and redistribution in the model with selfish voters may be overturned when one allows for heterogeneous preferences among the population. In particular, an increase in poverty is very likely to increase redistribution. Furthermore, this is also true for a mean-preserving spread leaving the median income unaffected, although it has no effect whatsoever on redistribution in the traditional selfish economy.

A few theoretical papers have recently introduced social preferences into the standard model of voting on redistribution. Dhami and Al-Nowaihi (2010a, b, and c) also use the RRMR framework in order to allow for fairness, using the utility function proposed by Fehr and Schmidt (1999), reflecting two-sided self-centered inequality aversion. This means that a voter dislikes both being poorer and richer than other voters, as he suffers disutility from any income difference between himself and the other individuals (i.e. there is both envy and altruism). They show that the existence of fair voters always increases the equilibrium level of redistribution as compared to an economy in which voters are self-interested. Furthermore, they show that in economies where the majority of voters are selfish, the decisive policy may be chosen by fair voters, and vice-versa. Finally, they demonstrate that increased poverty may lead to more redistribution in equilibrium, which is also true in our setup where selfish, rawlsian and utilitarian voters coexist.

Using another approach, Tyran and Sausgruber (2006) also assume inequity aversion *à la* Fehr and Schmidt (1999) in order to study voting on redistribution. They test their predictions, and find that the model with fair voters predicts voting outcomes far better than the standard model of voting assuming rationality and strict self-interest.

Galasso (2003) also introduces rawlsian altruism into Meltzer and Richard's (1981) framework, and allows for the coexistence of both selfish and rawlsian voters. However, he assumes away the possibility of some agents behaving partly as surplus maximizers (i.e. utilitarians), although experimental evidence indicates that they form a significant part of the population. Interestingly, it turns out that the presence of utilitarian voters does not alter the (qualitative) link between inequality and redistribution as compared with an economy composed exclusively of selfish and rawlsian voters.

Finally, Luttens and Valfort (2010) assume that voters care about others in a mix between maximizing the surplus (utilitarian concern) and helping the worst-off (rawlsian concern) (Charness and Rabin (2002)). They show that when altruistic preferences are desert-sensitive, that is, when there is a reluctance to redistribute from the hard-working to the lazy, the political equilibrium is characterized by lower levels of redistribution.

The rest of the paper is structured as follows: in Section 2, we describe the model, derive the preferred tax rates of the three types of voters, and characterize implicitly the Condorcet winner tax rate of the heterogeneous economy. Furthermore, we perform some simulations of the model in order to investigate further the properties of the equilibrium level of redistribution. In Section 3, we investigate the link between inequality and redistribution in the heterogeneous economy using three alternative definitions of inequality. In Section 4, we illustrate some implications of the model in the context of (the political economy of) border formation. Finally, Section 5 concludes.

2.2 The Model

We introduce heterogeneous social preferences in the standard RRMR framework, in which the proceeds from a linear tax are used to finance equal per-capita transfers to all voters.

There is a continuum of individuals of mass 1. The voters are differentiated by their ability level, which is also their wage rate, denoted by ω. Furthermore, the abilities are distributed in the interval $[\underline{\omega}, \overline{\omega}] \subset \mathbb{R}^+$ according to the continuous cumulative distribution function $F(.)$. The distribution of abilities has mean $\widetilde{\omega}$, and, consistently with empirical evidence, is assumed to be skewed, so that $\omega_m < \widetilde{\omega}$. A production function transforms labor into a consumption good, according to the worker's ability: $y(\omega) = \omega l(\omega)$, where $l(\omega)$ represents the amount of labor supplied by the agent with ability ω.

Each individual is endowed with a fixed time endowment of one unit and supplies l units of labor, where $0 \leq l \leq 1$. Hence, the wage rate offered to each worker-voter coincides with the marginal product (i.e. the skill level ω). Therefore, the gross tax income of a voter is given by $y = \omega l$, and its budget constraint by

$$0 \leq c \leq (1-t)y + b = (1-t)\omega l + b$$

where $t \in [0, 1]$ is the tax rate, and b is the uniform transfer given to each voter that equals the average tax proceeds, that is,

$$b = t\overline{y} = t \int_{\underline{\omega}}^{\overline{\omega}} \omega l(\omega) dF(\omega)$$

We consider a two-stage game. In the first stage, voters choose a tax rate, t, anticipating the outcome of the second stage. Consumers exhibit fairness by voting for the tax rate that would maximize social welfare as seen from their own perspective (see below). In the second stage, individuals choose their labor supply l so as to selfishly maximize their private utility. This determines the vector of labor supplies and indirect utilities.

2.2.1 Individual Choice of Labour Supply (Second-Stage Game)

Taking the redistributive policy of the government as given (i.e. t and b), labor supply is determined on the basis of private preferences. A voter has a private utility function $U(c, 1-l)$, over own consumption, c, and own leisure, $(1-l)$. All voters have the same private utility function, and thus they differ only in that they are endowed with different ability levels, ω. Following the literature, we assume that private utility is a quasi-linear function of the form[2]

$$U\left(c, 1-l\right) = c + u\left(1-l\right)$$

The optimization problem of a voter is given by

$$\underset{l}{Max}\ U\left(c, 1-l\right) \text{ such that } 0 \leq c \leq \left(1-t\right)\omega l + b$$

The economic optimization problem yields the usual result

$$\left(1-t\right)\omega = u'\left(1-l^*\right)$$

which implicitly defines l^* as a function of ω and t, and thus $y^* = \omega l^*$.

Following the literature, we assume that the quasi-linear utility function takes the following quadratic form in leisure:

$$U\left(c, 1-l\right) = c - \frac{1}{2}l^2$$

The utility function $U\left(c, 1-l\right)$ is twice differentiable, strictly concave in leisure, and the marginal utility of both consumption and leisure is positive:

- $\frac{\partial U(c,1-l)}{\partial c} = 1 > 0$
- $\frac{\partial U(c,1-l)}{\partial l} = -l < 0$
- $\frac{\partial^2 U(c(l),1-l)}{\partial l^2} = -1 < 0$

The first-order condition yields

$$l^* = \left(1-t\right)\omega$$

$$y^* = \omega l^* = \left(1-t\right)\omega^2$$

Hence, private preference satisfaction is measured by the indirect utility function υ:

[2]Note that assuming such linearity in consumption has an important implication regarding the utilitarian voters. When an individual's utility function is linear in c, maximizing the sum/average of utilities (i.e. utilitarianism) is independent of any distributional concerns. In contrast, if one assumes a utility function that is strictly concave in c, maximizing the sum of utilities requires perfect equality of consumption among individuals. Besides the fact that such (quasi)-linearity is convenient for the tractability of the analysis, it also reflects the assumption according to which utilitarianism may only entail efficiency concerns.

$$
\begin{aligned}
\upsilon &= (1-t)\omega l^* + b + u\left(l^*\right) \\
\Leftrightarrow \upsilon(t,b,\omega) &= \frac{1}{2}(1-t)^2\omega^2 + b \\
\Leftrightarrow \upsilon(t,\omega) &= \frac{1}{2}(1-t)^2\omega^2 + t(1-t)\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega)
\end{aligned}
$$

2.2.2 Voting for a Tax Rate (First-Stage Game)

All agents take their voting decisions by maximizing their indirect utility function. There is a proportion $\alpha \in (0,1)$ of the population that is selfish (S), and that maximizes

$$V^S(t,\omega) = \upsilon(t,\omega)$$

There is a proportion $\beta \in (0,1)$ of the population that is rawlsian (R), and that maximizes

$$V^R(t,\omega) = (1-\lambda^R)\upsilon(t,\omega) + \lambda^R \upsilon(t,\underline{\omega})$$

There is a proportion $\gamma \in (0,1)$ of the population that is utilitarian (U), and that maximizes

$$V^U(t,\omega) = (1-\lambda^U)\upsilon(t,\omega) + \lambda^U \int_{\underline{\omega}}^{\overline{\omega}} \upsilon(t,\omega) dF(\omega)$$

where $\lambda^i \in (0,1)$, $i = R, U$, are the relative weights associated to rawlsian and utilitarian motives, respectively. The indirect utility function of all types of voters is quasi-concave, and so preferences are single-peaked on the tax rate dimension. Proposition 1 below gives the preferred tax rate of each type of voter for a given ability level ω.

Proposition 1 (Preferred Tax Rates)**.**

1. *The preferred tax rate of a selfish voter with ability ω is given by*

$$t^S(\omega) = \frac{\overline{y} - y(\omega)}{2\overline{y} - y(\omega)}$$

2. *The preferred tax rate of a rawlsian voter with ability ω is given by*

$$t^R(\omega) = \frac{\overline{y} - \lambda^R y(\underline{\omega}) - (1-\lambda^R) y(\omega)}{2\overline{y} - \lambda^R y(\underline{\omega}) - (1-\lambda^R) y(\omega)}$$

3. *The preferred tax rate of a utilitarian voter with ability ω is given by*

$$t^U(\omega) = \frac{\overline{y} - \lambda^U \overline{y} - (1-\lambda^U) y(\omega)}{2\overline{y} - \lambda^U \overline{y} - (1-\lambda^U) y(\omega)}$$

Observe that selfish and utilitarian individuals never vote for a positive tax rate when their income is above average. In contrast, the rawlsian tax rate is strictly positive whenever $y(\omega) < \frac{\overline{y}-\lambda^R y(\underline{\omega})}{1-\lambda^R}$, which is strictly higher than $\overline{y}$ for $\lambda^R > 0$. Hence, even though a rawlsian individual is richer than average —and so does not benefit from redistribution himself— he may still vote for a positive level of redistribution, since he derives utility from giving away resources towards the poorest individual in the economy.

Proposition 2 (Tax Rates' Ordering)**.** *For a given ability level ω, the following relationship holds:*

$$t^U(\omega) \leqslant t^S(\omega) \leqslant t^R(\omega)$$

Clearly, for three distinct types of voters with an identical ability level ω, the one supporting the highest tax rate is rawlsian, while the one supporting the smallest tax rate is utilitarian (see Figure 2.1). As the rawlsian voter enjoys utility from redistributing to the poorest individual in the economy, he's also the one valuing redistribution the most. In contrast, utilitarian voters do not care about redistribution *per se*, and thus, as selfish individuals, they only vote for a positive tax rate to the extent that they benefit themselves from redistribution (i.e. $y(\omega) < \overline{y}$). However, since they are (partly) surplus maximizers, they support a strictly smaller tax rate than selfish voters for any ability level ω.

Proposition 3 below gives some comparative statics results regarding the preferred tax rate of the three types of voters.

Proposition 3 (Comparative Statics)**.** *Suppose that $t^i > 0$, $i = S, R, U$. Then it holds that:*

1. *The three tax rates are decreasing in own income $y(\omega)$, and increasing in average income $\overline{y}$*
2. *The rawlsian tax rate is increasing in λ^R*
3. *The utilitarian tax rate is decreasing in λ^U*

As for selfish voters (and thus as in the standard Meltzer and Richard's (1981) framework), rawlsian and utilitarian altruists vote for lower tax rates when their own income increases, and for higher tax rates when the average income increases. Indeed, if $y(\omega)$ decreases and/or $\overline{y}$ increases, a given voter with ability level ω gets relatively poorer, so that he prefers a strictly higher tax rate independently of his type.

Furthermore, rawlsian altruists vote for higher tax rates when the weight they associate to the maximin criterion increases, since they derive more utility from redistributing to the poorest individual. For a utilitarian/surplus maximizer voter, the opposite holds: the higher λ^U, the smaller his preferred tax rate. Indeed, when λ^U increases, (poorer-than-average) utilitarian voters care relatively more about minimizing the distortions associated to taxation, so that they correspondingly vote for smaller tax rates.

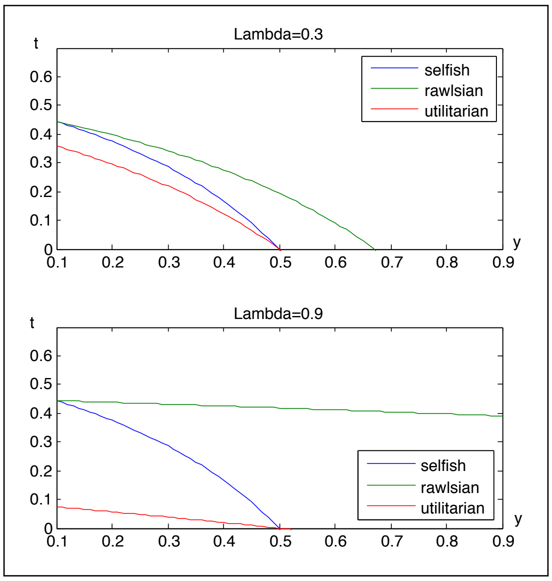

Figure 2.1: Tax rates versus income

Finally, observe that the higher the weights λ^U and λ^R, the bigger the difference between the preferred tax rates of the three types for a given ability level ω (see Figure 2.1).

The relationships between the incomes of the three types of individuals when they vote for the same tax rate t are the following:

$$y^S(t) = y^R(t) + \lambda^R \left[y\left(\underline{\omega}\right) - y^R(t)\right]$$

$$y^S(t) = y^U(t) + \lambda^U \left[\overline{y} - y^U(t)\right]$$

$$y^R(t) = \frac{(1-\lambda^U)}{(1-\lambda^R)} y^U(t) + \frac{\lambda^U}{(1-\lambda^R)}\overline{y} - \frac{\lambda^R}{(1-\lambda^R)} y\left(\underline{\omega}\right)$$

As can be seen in Figure 2.1, and as already noted, selfish and utilitarian individuals never vote for a positive tax rate when their income is above average. Furthermore, observe that a utilitarian individual votes at most for a tax rate of $\frac{1-\lambda^U}{2-\lambda^U}$. Therefore, we have the following lemma:

Lemma 1. *Assume that* $\lambda^R = \lambda^U = \lambda$*. If* $t^S = t^R = t^U = t \in \left(0, \frac{1-\lambda}{2-\lambda}\right)$*, then it holds that*

$$y^U(t) < y^S(t) < y^R(t)$$

Clearly, when the three types of voters support the same (positive) tax rate, the richest individual is rawlsian, while the poorest one is utilitarian. The utilitarian voter, as he cares about the maximization of the surplus, is consequently the unique poorer-than-average type bearing a cost from taxing the rich. Hence, as can be seen in Figure 2.1, when a utilitarian and a selfish voter have the same (below average) income, the selfish voter supports a strictly higher tax rate than the utilitarian one. We saw that the maximum value of the utilitarian tax rate is $\frac{1-\lambda}{2-\lambda}$. Clearly, this value is decreasing in the parameter λ. At the limit, when $\lambda = 0$, utilitarians become selfish, and hence vote at most for a tax rate of 1/2. Conversely, if $\lambda = 1$, that is, if utilitarians only care about surplus maximization, they are against any positive level of redistribution, so that $t^U(\omega) = 0$ for any ability level ω.

2.2.3 Political Equilibrium

In order to find the political equilibrium resulting from the interaction of selfish, rawlsian and utilitarian voters, we need to identify the Condorcet winner tax rate of this economy. In other words, we need to find the tax rate corresponding to the median, that is, such that half of the preferred tax rates are below it and the other half of the preferred tax rates are above it. In order to do so, we first express the three types' incomes as a function of their preferred tax rates:

$$y^S(t) = \frac{(1-2t)}{(1-t)}\overline{y}$$

$$y^R(t) = \frac{(1-2t)}{(1-t)(1-\lambda^R)}\overline{y} - \frac{\lambda^R}{(1-\lambda^R)}y(\underline{\omega})$$

$$y^U(t) = \frac{(1-2t)}{(1-t)(1-\lambda^U)}\overline{y} - \frac{\lambda^U}{(1-\lambda^U)}\overline{y}$$

As there is a one-to-one correspondence between the preferred tax rate and the income for each type, it is equivalent to identify the median income corresponding to the median tax rate. Therefore, the Condorcet winner tax rate is given by

$$t^* = \left\{ t | \alpha F\left(y^S(t)\right) + \beta F\left(y^R(t)\right) + \gamma F\left(y^U(t)\right) = \frac{1}{2} \right\}$$

$$\Leftrightarrow t^* = \left\{ t | \Phi(t) = \frac{1}{2} \right\}$$

In order to have an interior solution for the equilibrium tax rate of this economy, it must hold that

$$\begin{cases} \Phi(t) < \frac{1}{2} \text{ for some } t \in \left[0, \frac{1}{2}\right] \\ \Phi(t) > \frac{1}{2} \text{ for some } t \in \left[0, \frac{1}{2}\right] \\ \Phi(t) \text{ is continuous and monotonic on the interval } \left[0, \frac{1}{2}\right] \end{cases}$$

From the following lemma, it is direct that monotonicity is satisfied:

Lemma 2. $\Phi(t)$ *is strictly decreasing in* t *for all* $t < \frac{1}{2}$.

For the two extreme values of t, we have that

1. $y^S(0) = \overline{y}$, $y^R(0) = \frac{\overline{y} - \lambda^R y(\underline{\omega})}{1-\lambda^R} > \overline{y}$, and $y^U(0) = \overline{y} \Rightarrow \Phi(0) > 0$

2. $y^S\left(\frac{1}{2}\right) = 0$, $y^R\left(\frac{1}{2}\right) < 0$, and $y^U\left(\frac{1}{2}\right) < 0 \Rightarrow \Phi\left(\frac{1}{2}\right) = 0$

Hence, we have the following result:

Proposition 4 (Existence of the Political Equilibrium)**.**
If the distribution of abilities $F(\omega)$ is continuous and positively skewed, there exists a (unique) Condorcet winner tax rate $t^* > 0$, which is implicitly defined by $\Phi(t^*) = 1/2$.

Proof. Given that $y^i(t)$, $i = S, R, U$, are continuous in their domain, and $F(.)$ is continuous, $\Phi(t)$ is also continuous. Then, given that the distribution of abilities $F(\omega)$ is positively skewed, and that gross income $y = \omega l = (1-t)\omega^2$, it follows directly that the gross income distribution $F(y(\omega))$ is positively skewed as well, so that $y_m < \overline{y}$. Therefore, we have that $F(\overline{y}) > 1/2$, and thus

$$\Phi(0) = \alpha F(\overline{y}) + \beta F\left(\frac{\overline{y} - \lambda^R y(\underline{\omega})}{1-\lambda^R}\right) + \gamma F(\overline{y}) > \frac{1}{2}$$

Hence, we have that

1. $\Phi(0) > \frac{1}{2}$
2. $\Phi\left(\frac{1}{2}\right) = 0$
3. $\Phi(t)$ is continuous and strictly decreasing in t (Lemma 2)

and so there exists a unique value of t such that $\Phi(t) = 1/2$. Finally, any individual —independently of his type— with income $y(\omega) < \overline{y}$ always supports a tax rate that is strictly positive, from which it follows that $t^* > 0$.

□

From Lemma 1, we can identify which voter, within each type, votes for the Condorcet winner tax rate t^* of the economy (i.e. the decisive income, within each type):

$$y^U(t^*) < y^S(t^*) < y^R(t^*) \text{ for } t^* \in \left(0, \frac{1-\lambda}{2-\lambda}\right)$$

and

$$y^S(t^*) < y^R(t^*) \text{ for } t^* \in \left(\frac{1-\lambda}{2-\lambda}, \frac{1}{2}\right) \text{ and no utilitarian has } t^* \text{ as a preferred tax rate}$$

In order to perform comparative statics regarding the equilibrium value of the redistributive parameter, we do the following transformation on the proportion of each type, so that the effect of a marginal increase in a given proportion can be determined: the Condorcet winner tax rate is given by

$$t^* = \left\{ t | \alpha F\left(y^S(t)\right) + \beta F\left(y^R(t)\right) + \gamma F\left(y^U(t)\right) = \frac{1}{2} \right\}$$

Equivalently, we can write

$$t^* = \left\{ t | \frac{\alpha}{\alpha+\beta+\gamma} F\left(y^S(t)\right) + \frac{\beta}{\alpha+\beta+\gamma} F\left(y^R(t)\right) + \frac{\gamma}{\alpha+\beta+\gamma} F\left(y^U(t)\right) = \frac{1}{2} \right\}$$

Proposition 5 (Comparative Statics)**.** *For an economy composed of selfish, rawlsian and utilitarian voters with respective proportions α, β, and γ, the Condorcet winner tax rate is increasing in λ^R, and decreasing in λ^U. Furthermore, it is increasing in β, decreasing in γ, while the effect of an increase in α is ambiguous.*

As expected, the equilibrium tax rate is increasing in the intensity of rawlsian altruism, while it is decreasing in the intensity of utilitarian altruism. Furthermore, an increase in the proportion of rawlsian voters increases the equilibrium level of redistribution, whereas an increase in the proportion of utilitarian voters has the opposite effect. Finally, an increase in the relative proportion of selfish individuals has an ambiguous effect on the equilibrium tax rate. As we saw, the preferred tax rate of a selfish voter has an intermediate position between the rawlsian and the utilitarian tax rates, for a given ability level (Proposition 2). Furthermore, an increase in the relative proportion of selfish individuals translates into an equivalent decrease in the relative proportion of one of the other two types, or both. Hence, the total effect of an increase in α depends on the distance between the selfish tax rates and the utilitarian and rawlsian ones (and thus on λ^U and λ^R), as well as on how this change in α affects the two other remaining proportions (i.e. β and γ).

2.2.4 Simulations

In order to investigate further the properties of the equilibrium tax rate of this economy, we simulate the model assuming that individual (gross) income is lognormally distributed, that is,

$$y(\omega) \backsim \log N\left(\mu, \sigma^2\right)$$

where

$$mean\left(y(\omega)\right) = e^{\mu+\frac{\sigma^2}{2}},\ var\left(y(\omega)\right) = (e^{\sigma^2}-1)e^{2\mu+\sigma^2} \text{ and } med\left(y(\omega)\right) = e^{\mu}$$

The lognormal is a good approximation of empirical income distributions, leads to tractable results, and allows for an unambiguous definition of inequality (Benabou (2000)). We set $mean\left(y(\omega)\right) = 0.5$ and $var\left(y(\omega)\right) = 0.1$, so that the ratio of median-to-mean income is around 0.85.

Figure 2.2 depicts the Condorcet winner tax rate for $\lambda^i \in [0.1, 0.9]$, $i = R, U$, assuming the following relative proportions of selfish, rawlsian and utilitarian voters: $(\alpha, \beta, \gamma) = (0.44, 0.35, 0.21)$. These proportions are the ones that have been found experimentally by Andreoni and Miller (2002)[3].

[3]The experiments conducted by Fismal et al. (2007) and Iriberri and Rey-Biel (2008) give similar proportions.

As can be seen in the figure, t^* reaches its maximum value for $(\lambda^R, \lambda^U) = (0.9, 0.1)$ (i.e. $t^* = 0.3$), and it reaches its minimum value for $(\lambda^R, \lambda^U) = (0.1, 0.9)$ (i.e. $t^* = 0.05$). Therefore, even though the selfish individuals constitute the biggest group in the population (i.e. $\alpha = 0.44$), variations in the altruistic weights lead to very important corresponding changes in the redistribution level.

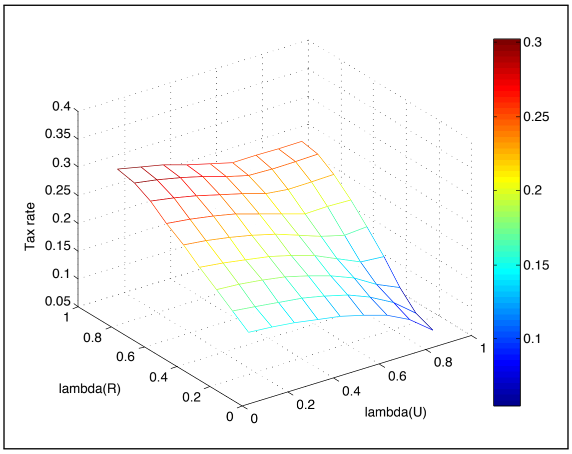

Figure 2.2: Condorcet winner tax rate as a function of the altruistic weights λ^R and λ^U

In Figure 2.3, we illustrate a rather extreme situation in which both utilitarian and rawlsian voters attach a very high relative weight to altruistic motives (i.e. $\lambda^U = \lambda^R = 0.9$). In this case, the equilibrium tax rate is equal to 0.41 when there are no utilitarian voters in the economy (i.e. $\alpha = \beta = 0.5$). Then, when the proportion of rawlsian voters decreases, the equilibrium tax rate decreases quickly to reach much lower levels, especially for low values of α. For instance, when $(\alpha, \beta, \gamma) = (0.1, 0.5, 0.4)$, the Condorcet winner tax rate of the economy is around 40%. Then, if the proportion of rawlsian voters decreases slightly to 0.35, with an equivalent increase in the proportion of utilitarian voters to 0.55, the equilibrium tax rate falls dramatically to reach a value below 5%.

Observe that for $\lambda^U = \lambda^R = 0.9$, the poorest utilitarian individual votes (approximatively) for $t = 0.08$. When $(\alpha, \beta, \gamma) = (0.1, 0.5, 0.4)$, the fact that $t^* = 0.4$ means that the proportion of both selfish and rawlsian voters who support $t^* > 0.4$ constitutes exactly one half of the population. Then, if the relative proportions change slightly to $(\alpha, \beta, \gamma) = (0.1, 0.35, 0.55)$, utilitarians are consequently needed in order to form a majority, and thus t^* has to decrease. In particular, t^* has to be smaller than 0.08, so that at least part of the utilitarians can be "attracted".

Therefore, it turns out that slight changes in the relative proportions of the respective types (selfish, rawlsian, utilitarian) can lead to very important variations in the extent of redistribution. Furthermore, this gets increasingly likely the bigger the altruistic weights in the voters' utility function.

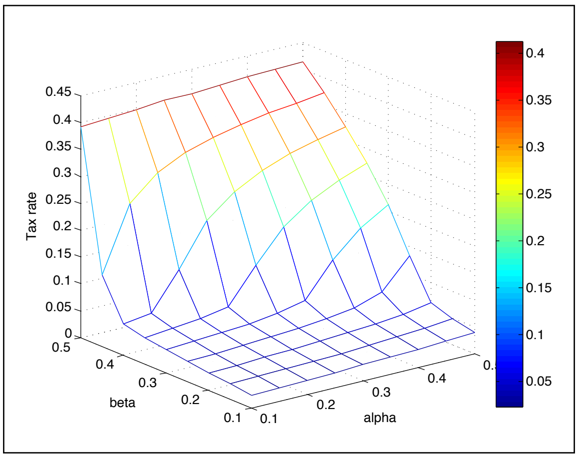

Figure 2.3: Condorcet winner tax rate as a function of the proportions α and β

2.3 The Link Between Inequality and Redistribution

The RRMR model predicts that higher inequality (lower median-to-mean income ratio) implies higher redistribution. Indeed, with selfish voters, an increase in inequality is relevant only to the extent that it concerns the relative position of the median voter. Empirically, this prediction remains controversial. Borck (2007, p. 96), in his survey on voting, inequality and redistribution, discusses this issue and concludes that " *[...] the RRMR hypothesis of a link between inequality and the size of the government has met with mixed empirical evidence*". In this section, we explore how the introduction of heterogeneous social preferences in the RRMR framework affects the predicted link between income inequality and redistribution.

We borrow the definition of inequality to Galasso (2003), who considers three types of income inequality: the poor getting poorer, the rich getting richer, and a mean-preserving spread.

Definition 1. *Let $F(.)$ be the original (positively skewed) cumulative distribution function of income, and let $G(.)$ be the (positvely skewed) cumulative distribution function of income after the change in inequality. Furthermore, let $y(\omega_1)$ and $y(\omega_2)$ be such that*

$$y(\underline{\omega}) < y(\omega_1) < \overline{y} < y(\omega_2) < y(\overline{\omega})$$

There are three types of income inequality:

1. *The poor get poorer if* $\begin{cases} F(y) = G(y) \text{ for } y > y(\omega_1) \\ F(y) \leq G(y) \text{ for } y \leq y(\omega_1) \end{cases}$ *and $y(\underline{\omega})$ and $\overline{y}$ decrease*

2. *The rich get richer if* $\begin{cases} F(y) = G(y) \text{ for } y < y(\omega_2) \\ F(y) \geq G(y) \text{ for } y \geq y(\omega_2) \end{cases}$ *and $y(\overline{\omega})$ and $\overline{y}$ increase*

3. *There is a mean-preserving spread if*

$$\begin{cases} F(y) \leq G(y) \text{ for } y \leq y(\omega_1) \\ F(y) = G(y) \text{ for } y(\omega_1) \leq y \leq y(\omega_2) \\ F(y) \geq G(y) \text{ for } y \geq y(\omega_2) \end{cases} \quad \text{and } y(\underline{\omega}) \text{ decreases, } y(\overline{\omega}) \text{ increases, with } \overline{y} \text{ unchanged}$$

In the traditional RRMR framework with selfish voters, and given that $y_m < \overline{y}$, the equilibrium level of redistribution always increases when the rich get richer (according to the above definition), since the median voter is relatively poorer. Then, a mean-preserving spread leaving the median income unaffected (i.e. $y_m > y(\omega_1)$) has no effect on the equilibrium tax rate. Finally, if the poor get poorer, the equilibrium level of redistribution decreases provided that $y_m > y(\omega_1)$, since the median voter is relatively richer (i.e. $\overline{y}$ decreases).

The following proposition describes the effect of an increase in the three types of inequality on the Condorcet winner tax rate of the economy, and shows that it yields different predictions than for the case of purely self-interested voters.

Proposition 6. *Let t^* be the equilibrium tax rate under $F(.)$, and let t^{**} be the equilibrium tax rate under $G(.)$. In an economy with selfish, rawlsian and utilitarian voters with respective proportions α, β, and γ,*

1. *If the poor get poorer, the effect on the equilibrium tax rate is ambiguous*
2. *If the rich get richer and $y^R(t^*) < y(\omega_2)$, then $t^{**} > t^*$*
3. *If there is a mean-preserving spread and $y^R(t^*) < y(\omega_2)$, then $t^{**} > t^*$*

If the rich get richer, and provided that $y^R(t^*) < y(\omega_2)$, none of the decisive voters within each type is directly affected by the rise in inequality. Therefore, in that case, the only relevant effect is the corresponding increase in the average income $\overline{y}$, which translates into a bigger mass of voters supporting $t > t^*$, so that t^* increases.

If there is a mean-preserving spread and $y^S(t^*) \leqslant y(\omega_1)$, the mass of selfish voters who supports $t > t^*$ increases, while it remains the same if $y^S(t^*) > y(\omega_1)$, since the average income $\overline{y}$ is unaffected by the rise in inequality. The same is true for utilitarians. Then, if $y^R(t^*) < y(\omega_2)$, although the rawlsian decisive voter is not affected directly by the rise in inequality, the mass of rawlsian voters supporting $t > t^*$ increases, since the income of the poorest individual has decreased and $\frac{\partial y^R(t^*)}{\partial y(\underline{\omega})} < 0$. Therefore, when there is a mean-preserving spread and $y^R(t^*) < y(\omega_2)$, t^* has to increase.

If the poor get poorer, things are slightly more complex. If $y^S(t^*) > y(\omega_1)$, the selfish decisive voter is not affected directly by the rise in inequality, and the only relevant effect is the corresponding decrease in the average income $\overline{y}$, so that the mass of selfish voters supporting $t > t^*$ decreases. Conversely, if $y^S(t^*) \leqslant y(\omega_1)$, the effect is ambiguous, since besides the decrease in $\overline{y}$, the selfish decisive voter is now poorer (in absolute terms), which makes him value strictly more redistribution.

The same is true for utilitarians. Then, if $y^R(t^*) > y(\omega_1)$, the rawlsian decisive voter may prefer either a lower or a higher tax rate, since the effect of the decrease in $y(\underline{\omega})$ and $\overline{y}$ go in opposite directions. The same is true if $y^R(t^*) \leqslant y(\omega_1)$, although in this case the positive effect on t^* is reinforced, since the decisive rawlsian voter is now poorer (in absolute terms). Hence, altogether, the effect on the equilibrium tax rate t^* when the poor get poorer is ambiguous, since the mass of voters supporting $t > t^*$ may either increase or decrease.

2.4 An Application to the Political Economy of Border Formation

Bolton and Roland (1997) have shown that income-based redistribution has two effects on the incentives for a given region to secede from a union: a political effect, as the regional and national median incomes differ, and a tax base effect, as average income differs between regions. The political effect reflects differences in preferences for redistribution, and always induces a given region to secede, independently of the existence of interregional transfers. Such transfers arise when regional average incomes differ, and typically induce richer regions to secede (the tax-base effect).

In this section, we illustrate with a few examples the fact that even though there are no income differences whatsoever between regions (and thus no political nor tax-base effect), a region may still prefer to secede from a union provided that its distribution of types (i.e. selfish, rawlsian, utilitarian) is different than the one of the union. Furthermore, when there are income differences between regions regarding their average and/or median income, this "type distribution" effect then interacts with the two above-mentioned "income distribution" effects in non-trivial ways so as to shape the regional incentives to (voluntarily) form part of a political union.

Assume there are three regions A, B, C, and three income types (i.e. individulas) in each region given by $y_1^i < y_2^i < y_3^i$, $i = A, B, C$. Assume, furthermore, that the (gross) income distribution in each region is positively skewed, that is,

$$y_m^i = y_2^i < \frac{(y_1^i + y_2^i + y_3^i)}{3} = \overline{y}^i$$

so that there is a positive level of redistribution in all the three regions when they are independent. Finally, assume that individuals can be either selfish (S), rawlsian (R), or utilitarian (U) (with $\lambda^U = \lambda^R = \lambda$), and let the preferred tax rate of a type k individual ($k = S, R, U$) with income y_j ($j = 1, 2, 3$) in region i ($i = A, B, C$) be given by

$$t^k\left(y_j^i\right)$$

Example 1 (Selfish Economy - No Income Differences Between Regions). *Suppose* $y_j^A = y_j^B = y_j^C = y_j$, $j = 1, 2, 3$. *That is, there are no income differences between regions. Suppose, furthermore, that all individuals are selfish. In this case, in each region, it holds that* $t^S(y_1) > t^S(y_2) > t^S(y_3) = 0$, *and* $t^S(y_2)$ *is implemented in each one of them if there are independent. If the three regions unify, then, the median tax rate is the one preferred by the median-income class, and thus coincides with the* *equilibrium tax rate in each region under independence. Since all regions are equally rich, and given that the median income is the same across regions, all voters are indifferent between independence and unification.*

Example 2 (Selfish and Utilitarian Economies - No Income Differences Between Regions). *Suppose now that $y_j^A = y_j^B = y_j^C = y_j$, $j = 1, 2, 3$, but individuals in region A are utilitarian altruists, while individuals in regions B and C are selfish. In this case, we have the following ordering of the preferred tax rates:*

$$t^S(y_1) > t^S(y_2) > t^U(y_2) > t^S(y_3) = t^U(y_3) = 0$$

and

$$t^U(y_1) > t^S(y_2) \text{ if and only if } y_2 - (1-\lambda)y_1 > \lambda\overline{y}$$

If the three regions are independent, the median-income voter is decisive in each one of them. Now, if the three regions unify, the median tax rate is either $t^U(y_1)$ or $t^S(y_2)$, depending on which one of the two is higher. Suppose that $t^U(y_1) > t^S(y_2)$. In that case, the equilibrium (median) tax rate is $t^S(y_2)$, from which it follows that all individuals in region B and C are indifferent between independence and unification. However, A majority of individuals in region A prefer a tax rate strictly lower than $t^S(y_2)$, and so region A does not join the union if integration is voluntary.

Suppose now that $t^U(y_1) < t^S(y_2)$. In that case, the equilibrium (median) tax rate is $t^U(y_1)$, and thus the lowest income is now decisive in setting the level of redistribution in the unified country. In that case, a majority of voters in region A prefer a tax rate strictly lower than $t^U(y_1)$, while a majority of individuals in region B and C prefer a tax rate strictly higher than $t^U(y_1)$. Therefore, although there are no income differences whatsoever between the three regions, none of them is willing to form a union.

Example 3 (Heterogeneous economies - No Income Differences Between Regions). *Suppose now that $y_j^A = y_j^B = y_j^C = y_j$, $j = 1, 2, 3$, and we have the following distribution of types in each region: in region A, the lowest and highest-income individuals are rawlsian, while the median-income one is selfish, that is, we have (R, S, R) in region A. Furthermore, suppose that we have (U, U, R) in region B, and (R, U, U) in region C. In this case, we have the following ordering of the preferred tax rates:*

$$t^R(y_1) > t^S(y_2) > t^U(y_2) > t^U(y_3) = 0$$

As in the previous example, $t^U(y_1)$ may be either lower or higher than $t^S(y_2)$, while $t^R(y_3)$ may have any position between $t^R(y_1)$ and $t^U(y_3)$. Suppose that we have the following ordering:

$$t^R(y_1) > t^R(y_3) > t^S(y_2) > t^U(y_1) > t^U(y_2) > t^U(y_3) = 0$$

In that case, we have that $t^R(y_3)$, $t^U(y_1)$, and $t^U(y_2)$ are respectively implemented in region A, B and C when they are independent. Observe that in region A, although a majority of voters are selfish, the redistributive outcome is controlled by a rawlsian voter. Notice, furthermore, than thanks to the presence of the rawlsian high-income voter in region B, the low-income utilitarian voter is decisive in setting the tax rate. Therefore, in both regions A and B, the coexistence of different types of voters reduces the decisiveness of the median-income class.

If the three regions unify, the median tax rate is then given by $t^S(y_2)$. Observe that although the selfish voters form a minority in the unified country (i.e. 1/9), it is a selfish individual that controls the redistributive outcome. In region A, a majority of individuals prefer $t > t^S(y_2)$, while a majority of individuals in region B and C prefer $t < t^S(y_2)$. Again, although there are no income differences across regions, none of them is willing to form a union.

Example 4 (Heterogeneous economies - Income Differences Between Regions)**.** *Suppose now that there are only 2 regions A and B, and assume the following distribution of types in each region: (S, R, R) in region A, and (S, R, U) in region B. Suppose, furthermore, that $y_1^A > y_1^B$, $y_2^A = y_2^B$, and $y_3^A < y_3^B$. Finally, assume that the distribution of income in B is a mean-preserving spread of the one in A, so that $\overline{y}^A = \overline{y}^B$. Under **independence**, we have the following ordering of the tax rates:*

$$t^S(y_1^B) > t^S(y_1^A) > t^R(y_2^B) > t^R(y_2^A) > t^R(y_3^A) \geqslant t^U(y_3^B) = 0$$

and thus, if A and B are independent, the median tax rates are given by $t^R(y_2^A)$ and $t^R(y_2^B)$ respectively. Observe that even though the median-income (decisive) voter in both regions is of the same type with the same income, $t^R(y_2^B) > t^R(y_2^A)$ since $y_1^A > y_1^B$.

*Under **unification** of the two regions, we have the following ordering of the tax rates:*

$$t^S(y_1^B) > t^S(y_1^A) > t^R(y_2^B) = t^R(y_2^A) > t^R(y_3^A) \geqslant t^U(y_3^B) = 0$$

The rawlsian voter in A and B now have the same preferred tax rate (since the reference income for both of them is y_1^B), which is decisive under unification. In region B, all voters are indifferent between unification and independence, since the level of redistribution is unaffected, and there are no interregional transfers taking place under unification (recall that $\overline{y}^A = \overline{y}^B$). In region A, since unification yields more redistribution, it follows that the poorest (selfish) individual is unambiguously better off under unification. Then, the median and high-income rawlsian individuals may be better off under unification if they care enough about helping the poor, that is, if λ is high enough.

Hence, differences in income and type distributions now interact to shape the incentives to unify in each region. If, in addition to that, $\overline{y}^A \neq \overline{y}^B$, there is an additional tax-base effect that will, all other things being equal, induce the richer region to prefer independence, and the poorer region to prefer unification.

2.5 Conclusion

We endowed individuals with heterogeneous social preferences in the traditional RRMR framework on voting on redistribution. More specifically, we assumed that selfish, rawlsian and utilitarian voters coexist with given proportions. We characterized implicitly the Condorcet winner tax rate of this economy, and proved its existence for any positively skewed distribution of income. While an increase in the relative proportion of rawlsian (utilitarian) voters always yields more (less) redistribution in equilibrium, the effect of an increase in the relative proportion of selfish voters has an ambiguous effect on the equilibrium tax rate. In fact, given that the preferred tax rates of selfish voters have an intermediate position between the ones of rawlsian and utilitarian voters, the equilibrium level of redistribution in the heterogeneous economy may be either lower or higher than the one in an economy composed exclusively of selfish voters.

By simulating the model, we showed that small variations in the relative proportions of the three types may have a very large impact on the extent of redistribution when rawlsian and utilitarian voters attach a high relative weight to altruistic motives in their utility function.

Regarding the link between inequality and redistribution, it turns out that allowing for the coexistence of selfish, rawlsian and utilitarian voters yields slightly different predictions than when all voters are purely self-interested. In particular, an increase in poverty may increase the equilibrium level of redistribution. Furthermore, in case of a mean-preserving spread leaving the median income unaffected, which has no effect on redistribution in a selfish economy, redistribution is likely to increase in the heterogeneous economy, since the decrease in the lowest income induces the rawlsian voters to vote for more redistribution, and, in addition to that, the income position of the decisive voter within each type does not necessarily correspond to the median. Finally, if the rich get richer, the effect on redistribution is the same (qualitatively) as in the case of selfish voters, provided that the rawlsian decisive voter is not directly affected by the rise in inequality.

We assumed heterogeneity with respect to the *notion* of fairness individuals include in their preferences. However, a possible extension could be to consider heterogeneity with respect to the *weight* people attach to fairness considerations. Cappelen et al. (2007) showed that "*[...] both kinds of heterogeneity matter in explaining individual behavior*". However, observe that this requires making additional assumptions regarding the distribution of fairness weights among the voters. In particular, we would like to know how the fairness weights relate to income. If they are positively related, the relationship between the preferred tax rate and income will not anymore be monotonic for the rawlsian voters. If they are negatively related, the same holds for the utilitarian voters. This obviously has implications for the resulting equilibrium. Similarly, it could also well be the case that fairness intensity, rather than being related to income, changes according to whom redistribution applies to (e.g. immigrants). Finally, the fairness weights could also evolve over time, or change according to the extent of inequality, or change according to the information available. Indeed, Iriberri and Rey-Biel (2008) have shown experimentally that knowing the distribution of types among the population might change how individuals feel about others. In particular, subjects that exhibit other-regarding preferences tend to behave more selfishly once they are provided with social information.

More generally, this raises the question of why people exhibit social preferences in the first place. In other words, it raises the issue of the endogeneity of fairness preferences in the model, which is beyond the scope of this paper. What we showed in this paper is that fairness, both in type and intensity, does matter for the equilibrium level of redistribution. Therefore, understanding who behaves fairly and why, as well as with which intensity, seems worth investigating.

2.6 References

1. Andreoni, J. and Miller, J. (2002), Giving According to GARP: An Experimental Test of the Consistency of Preferences for Altruism, *Econometrica*, Vol. 70, No. 2, pp. 737-753.

2. Ackert, L. F., Gillette, A. B., Martinez-Vasquez, J. and Rider, M. (2007a), Voting on Tax Policy Design: A Test of the Selfish versus Social Preferences Hypotheses, *Public Finance Review*, Vol. 35, No.2, pp. 263-284.

3. Ackert, L. F., Martinez-Vasquez, J. and Rider, M. (2007b), Social Preferences and Tax Policy Design: Some Experimental Evidence, *Economic Inquiry*, Vol. 45, No. 3, pp. 487-501.

4. Benabou, R. (2000), Unequal Societies: Income Distribution and the Social Contract, *American Economic Review*, Vol. 90, No. 1, pp. 96-129.

5. Bolton, G. E. and Ockenfels, A. (2002), The Behavioral Tradeoff between Efficiency and Equity when a Majority Rules, Discussion Papers on Strategic Interaction from Max Planck Institute of Economics, Strategic Interaction Group.

6. Bolton, P. and Roland, G. (1997), The Breakup of Nations: A Political Economy Analysis, *Quarterly Journal of Economics*, Vol. 112, No. 4, pp. 1057-1090.

7. Borck, R. (2007), Voting, Inequality and Redistribution, *Journal of Economic Surveys*, Vol. 21, No. 1, pp. 90-109.

8. Cappelen, A.W., Hole, A.D., Sorensen, E.Ø. and Tungodden, B. (2007), The Pluralism of Fairness Ideals: an Experimental Approach, *American Economic Review*, Vol. 97, No. 3, pp. 818-827.

9. Charness, G and Rabin, M. (2002), Understanding Social Preferences with Simple Tests, *Quarterly Journal of Economics*, Vol. 117, No. 3, pp. 817–869.

10. Dawes, C. T., Loewen, P. J. and Fowler, J. H. (2008), Social Preferences and Political Participation, Working Paper.

11. Dhami, S. and al-Nowaihi, A. (2010a), Redistributive Policy with Heterogenous Social Preferences of Voters, *European Economic Review*, Vol. 54, Issue 6, pp. 743759.

12. Dhami, S. and al-Nowaihi, A. (2010b), Existence of a Condorcet Winner When Voters Have Other Regarding Preferences, *Journal of Public Economic Theory*, Vol. 12, Issue 5, pp. 897-922.

13. Dhami, S. and al-Nowaihi, A. (2010c), Inequality and Redistribution When Voters Have Other Regarding Preferences, Unpublished manuscript, University of Leicester.

14. Edlin, A., Gelman, A. and Kaplan, N. (2007), Voting as a Rational Choice: Why and How People Vote to Improve the Well-Being of Others, *Rationality and Society*, Vol. 19, No. 3, pp. 293-314.

15. Erlei, M. (2008), Heterogeneous Social Preferences, *Journal of Economic Behavior and Organization*, Vol. 65, Issues 3-4, pp. 436-457.

16. Fehr, E. and Schmidt, K. (1999), A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics*, Vol. 114, No. 3, pp. 817-868.

17. Fehr, E. and Schmidt, K. (2006), The Economics of Reciprocity, Fairness and Altruism - Experimental Evidence and New Theories, in *Handbook of the Economics of Giving, Altruism and Reciprocity*, pp. 615–692, North-Holland.

18. Fisman, R., Kariv, S. and Markovitz, D. (2007), Individual Preferences for Giving, *American Economic Review*, Vol. 97, No. 2, pp. 153-158.

19. Galasso, V. (2003), Redistribution and Fairness: A Note, *European Journal of Political Economy*, Vol. 19, No. 4, pp 885-892.

20. Iriberri, N. and Rey-Biel, P. (2008), Elicited Beliefs and Social Information in Modified Dictator Games: What Do Dictators Believe Other Dictators Do?, Working Paper No. 1137, Universitat Pompeu Fabra.

21. Luttens, R. I. and M-A. Valfort (2010), Voting for Redistribution under Desert-Sensitive Altruism, *Scandinavian Journal of Economics*, forthcoming.

22. Meltzer, A. H. and Richard, S.F. (1981), A Rational Theory of the Size of Government, *Journal of Political Economy*, Vol. 89, No. 5, pp. 914–927.

23. Roberts (1977), Voting over Income Tax Schedules, *Journal of Public Economics*, Vol. 8, Issue 3, pp. 329-340.

24. Romer (1975), Individual Welfare, Majority Voting, and the Properties of a Linear Income Tax, *Journal of Public Economics*, Vol. 4, Issue 2, pp. 163-185.

25. Tyran, J-R. and Sausgruber, R. (2006), A Little Fairness May Induce a Lot of Redistribution in Democracy, *European Economic Review,* Vol. 50, No. 2, pp. 469-485.

2.7 Appendix

Proof of Proposition 1.

1. For a **selfish** individual with ability ω, we have

$$V^S(t,\omega) = v(t,\omega) = \frac{1}{2}(1-t)^2\omega^2 + t(1-t)\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega)$$

Taking partial derivative with respect to the tax rate t,

$$\frac{\partial V^S(t,\omega)}{\partial t} = -(1-t)\omega^2 + (1-2t)\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega)$$

Setting this quantity equal to zero and solving for t yields

$$t^S = \frac{\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega) - \omega^2}{2\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega) - \omega 2} = \frac{\overline{y} - y(\omega)}{2\overline{y} - y(\omega)}$$

2. For a **rawlsian** individual with ability ω, we have

$$V^R(t,\omega) = (1-\lambda^R)\left[\frac{1}{2}(1-t)^2\omega^2\right] + \lambda^R\left[\frac{1}{2}(1-t)^2\underline{\omega}^2\right] + t(1-t)\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega)$$

Taking partial derivative with respect to the tax rate t,

$$\frac{\partial V^R(t,\omega)}{\partial t} = -(1-\lambda^R)(1-t)\omega^2 - \lambda^R\underline{\omega}^2(1-t) + (1-2t)\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega)$$

Setting this quantity equal to zero and solving for t yields

$$t^R = \frac{\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega) - \lambda^R\underline{\omega}^2 - (1-\lambda^R)\omega^2}{2\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega) - \lambda^R\underline{\omega}^2 - (1-\lambda^R)\omega^2} = \frac{\overline{y} - \lambda^R y(\underline{\omega}) - (1-\lambda^R)y(\omega)}{2\overline{y} - \lambda^R y(\underline{\omega}) - (1-\lambda^R)y(\omega)}$$

3. For a **utilitarian** individual with ability ω, we have

$$V^U(t,\omega) = (1-\lambda^U)\left[\frac{1}{2}(1-t)^2\omega^2 + t(1-t)\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega)\right]$$

$$+\lambda^U\left[\frac{1}{2}(1-t)^2\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega) + t(1-t)\int_{\underline{\omega}}^{\overline{\omega}}\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega)dF(\omega)\right]$$

Taking partial derivative with respect to the tax rate t,

$$\frac{\partial V^U(t,\omega)}{\partial t} = -(1-t)(1-\lambda^U)\omega^2 + (1-\lambda^U)(1-2t)\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega)$$

$$-\lambda^U(1-t)\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega) + \lambda^U(1-2t)\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega)$$

Setting this quantity equal to zero and solving for t yields

$$t^U = \frac{\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega) - \lambda^U \int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega) - (1-\lambda^U)\omega^2}{2\int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega) - \lambda^U \int_{\underline{\omega}}^{\overline{\omega}} \omega^2 dF(\omega) - (1-\lambda^U)\omega^2} = \frac{\overline{y} - \lambda^U\overline{y} - (1-\lambda^U)y(\omega)}{2\overline{y} - \lambda^U\overline{y} - (1-\lambda^U)y(\omega)}$$

□

Proof of Proposition 2. It follows directly from the analytical expression of t^S, t^R and t^U that

1. If $y(\omega) < \overline{y}$, it holds that $0 < t^U(\omega) < t^S(\omega) < t^R(\omega)$
2. If $\overline{y} < y(\omega) < \frac{\overline{y}-\lambda^R y(\underline{\omega})}{1-\lambda^R}$, it holds that $0 = t^U(\omega) = t^S(\omega) < t^R(\omega)$
3. If $\frac{\overline{y}-\lambda^R y(\underline{\omega})}{1-\lambda^R} < y(\omega)$, it holds that $0 = t^U(\omega) = t^S(\omega) = t^R(\omega)$

Therefore, for any ability level ω, it holds that $t^U(\omega) \leqslant t^S(\omega) \leqslant t^R(\omega)$. □

Proof of Proposition 3. Suppose that $y(\omega) < \overline{y}$, so that $t^i > 0$, $i = S, R, U$.

1. For a **selfish** individual, we have

$$t^S = \frac{\overline{y} - y(\omega)}{2\overline{y} - y(\omega)}$$

Taking derivatives,

$$\frac{\partial t^S}{\partial y(\omega)} = -\frac{\overline{y}}{[2\overline{y} - y(\omega)]^2} < 0$$

$$\frac{\partial t^S}{\partial \overline{y}} = \frac{y(\omega)}{[2\overline{y} - y(\omega)]^2} > 0$$

2. For a **rawlsian** individual, we have

$$t^R = \frac{\overline{y} - \lambda^R y(\underline{\omega}) - (1-\lambda^R)y(\omega)}{2\overline{y} - \lambda^R y(\underline{\omega}) - (1-\lambda^R)y(\omega)}$$

Taking derivatives,

$$\frac{\partial t^R}{\partial y(\omega)} = -\frac{(1-\lambda^R)\overline{y}}{\left[2\overline{y} - \lambda^R y(\underline{\omega}) - (1-\lambda^R) y(\omega)\right]^2} < 0$$

$$\frac{\partial t^R}{\partial \overline{y}} = \frac{\lambda^R y(\underline{\omega}) + (1-\lambda^R) y(\omega)}{\left[2\overline{y} - \lambda^R y(\underline{\omega}) - (1-\lambda^R) y(\omega)\right]^2} > 0$$

$$\frac{\partial t^R}{\partial \lambda^R} = \frac{\overline{y}\left[y(\omega) - y(\underline{\omega})\right]}{\left[2\overline{y} - \lambda^R y(\underline{\omega}) - (1-\lambda^R) y(\omega)\right]^2} > 0$$

3. For a **utilitarian** individual, we have

$$t^U = \frac{\overline{y} - \lambda^U \overline{y} - (1-\lambda^U) y(\omega)}{2\overline{y} - \lambda^U \overline{y} - (1-\lambda^U) y(\omega)}$$

Taking derivatives,

$$\frac{\partial t^U}{\partial y(\omega)} = -\frac{(1-\lambda^U)\overline{y}}{\left[2\overline{y} - \lambda^U \overline{y} - (1-\lambda^U) y(\omega)\right]^2} < 0$$

$$\frac{\partial t^U}{\partial \overline{y}} = \frac{\lambda^U \overline{y} + (1-\lambda^U) y(\omega)}{\left[2\overline{y} - \lambda^U \overline{y} - (1-\lambda^U) y(\omega)\right]^2} > 0$$

$$\frac{\partial t^U}{\partial \lambda^U} = \frac{\left[y(\omega) - \overline{y}\right]\overline{y}}{\left[2\overline{y} - (1-\lambda^U) y(\omega) - \lambda^U \overline{y}\right]^2} < 0$$

□

Proof of Lemma 1.

1. Selfish versus Rawlsian: $t^S = t^R = t > 0$ is equivalent to

$$y^S(t) = \lambda^R y\left(\underline{\omega}\right) + (1-\lambda^R) y^R(t) = y^R(t) + \lambda^R\left[y\left(\underline{\omega}\right) - y^R(t)\right] < y^R(t)$$

Hence, if a selfish and fair rawlsian voter support the same positive tax rate, the rawlsian agent has a higher income than the selfish one.

2. Selfish versus Utilitarian: $t^S = t^U = t > 0$ is equivalent to

$$y^S(t) = (1-\lambda^U) y^U(t) + \lambda^U \overline{y} = y^U(t) + \lambda^U\left[\overline{y} - y^U(t)\right] > y^U(t)$$

Hence, if a selfish and utilitarian voter support the same positive tax rate, the selfish agent has a higher income than the utilitarian one.

3. Rawlsian versus Utilitarian: $t^R = t^U > 0$ is equivalent to

$$y^R(t) = \frac{(1-\lambda^U)}{(1-\lambda^R)} y^U(t) + \frac{\lambda^U}{(1-\lambda^R)}\overline{y} - \frac{\lambda^R}{(1-\lambda^R)} y\left(\underline{\omega}\right)$$

For $\lambda^R = \lambda^U = \lambda$, this boils down to

$$y^R(t) = y^U(t) + \frac{\lambda}{(1-\lambda)}\left[\overline{y} - y\left(\underline{\omega}\right)\right] > y^U(t)$$

Hence, if a utilitarian and fair rawlsian voter support the same positive tax rate, the rawlsian agent has a higher income than the utilitarian one.

□

Proof of Lemma 2.

$$\frac{\partial \Phi(t)}{\partial t} = \alpha f\left(y^S(t)\right)\frac{\partial y^S(t)}{\partial t} + \beta f\left(y^R(t)\right)\frac{\partial y^R(t)}{\partial t} + \gamma f\left(y^U(t)\right)\frac{\partial y^U(t)}{\partial t}$$

and thus

$$\frac{\partial \Phi\left(t\right)}{\partial t} < 0 \text{ for all } t < \frac{1}{2}$$

□

Proof of Proposition 5. The function $\Phi(t)$ is given by

$$\Phi(t) = \alpha F\left(y^S(t)\right) + \beta F\left(y^R(t)\right) + \gamma F\left(y^U(t)\right)$$

Substituting yields

$$\Phi(t) = \alpha F\left(\tfrac{(1-2t)}{(1-t)}\overline{y}\right) + \beta F\left(\tfrac{(1-2t)}{(1-t)(1-\lambda^R)}\overline{y} - \tfrac{\lambda^R}{(1-\lambda^R)}y(\underline{\omega})\right) + \gamma F\left(\tfrac{(1-2t)}{(1-t)(1-\lambda^U)}\overline{y} - \tfrac{\lambda^U}{(1-\lambda^U)}\overline{y}\right)$$

Taking derivatives,

$$\frac{\partial \Phi}{\partial \lambda^R} = \beta f\left(y^R(t)\right)\left[\frac{(1-2t)\overline{y} - (1-t)y(\underline{\omega})}{(1-t)(1-\lambda^R)^2}\right] > 0$$

$$\frac{\partial \Phi}{\partial \lambda^U} = -\gamma f\left(y^U(t)\right)\left[\frac{t\overline{y}}{(1-t)(1-\lambda^U)^2}\right] < 0$$

The Condorcet winner tax rate is given by

$$t^* = \left\{t|\alpha F\left(y^S(t)\right) + \beta F\left(y^R(t)\right) + \gamma F\left(y^U(t)\right) = \frac{1}{2}\right\}$$

Equivalently, we can write

$$t^* = \left\{ t | \frac{\alpha}{\alpha+\beta+\gamma} F\left(y^S(t)\right) + \frac{\beta}{\alpha+\beta+\gamma} F\left(y^R(t)\right) + \frac{\gamma}{\alpha+\beta+\gamma} F\left(y^U(t)\right) = \frac{1}{2} \right\}$$

This way, we can determine the effect on the Condorcet winner tax rate of a marginal increase in the proportion of each type.

We know that, when the three types vote for the same tax rate t, the following relationship holds:

$$y^R(t) > y^S(t) > y^U(t)$$

and thus

$$F\left(y^R(t)\right) > F\left(y^S(t)\right) > F\left(y^U(t)\right)$$

Therefore, taking derivatives, we get

$$\frac{\partial \Phi}{\partial \alpha} = \frac{1}{(\alpha+\beta+\gamma)^2} \left\{ \beta \left[F\left(y^S(t)\right) - F\left(y^R(t)\right) \right] + \gamma \left[F\left(y^S(t)\right) - F\left(y^U(t)\right) \right] \right\} \lesseqgtr 0$$

$$\frac{\partial \Phi}{\partial \beta} = \frac{1}{(\alpha+\beta+\gamma)^2} \left\{ \alpha \left[F\left(y^R(t)\right) - F\left(y^S(t)\right) \right] + \gamma \left[F\left(y^R(t)\right) - F\left(y^U(t)\right) \right] \right\} > 0$$

$$\frac{\partial \Phi}{\partial \gamma} = \frac{1}{(\alpha+\beta+\gamma)^2} \left\{ \alpha \left[F\left(y^U(t)\right) - F\left(y^S(t)\right) \right] + \beta \left[F\left(y^U(t)\right) - F\left(y^R(t)\right) \right] \right\} < 0$$

Using the implicit function theorem, we finally get the following comparative statics results regarding the Condorcet winner tax rate of an economy with heterogeneous social preferences:

$$\frac{\partial t}{\partial \lambda^R} > 0 \text{ and } \frac{\partial t}{\partial \lambda^U} < 0$$

$$\frac{\partial t}{\partial \alpha} \lesseqgtr 0 \text{ , } \frac{\partial t}{\partial \beta} > 0 \text{ and } \frac{\partial t}{\partial \gamma} < 0$$

□

Proof of Proposition 6. Let $y(\omega_1)$ and $y(\omega_2)$ be such that

$$y(\underline{\omega}) < y(\omega_1) < \overline{y} < y(\omega_2) < y(\overline{\omega})$$

Furthermore, recall that

$$y^S(t^*) = \frac{(1-2t^*)}{(1-t^*)} \overline{y}$$

$$y^R(t^*) = \frac{(1-2t^*)}{(1-t^*)(1-\lambda^R)} \overline{y} - \frac{\lambda^R}{(1-\lambda^R)} y(\underline{\omega})$$

$$y^U(t^*) = \frac{(1-2t^*)}{(1-t^*)(1-\lambda^U)}\overline{y} - \frac{\lambda^U}{(1-\lambda^U)}\overline{y}$$

From Lemma 1, we know that

$$y^U(t^*) < y^S(t^*) < y^R(t^*) \text{ for } t^* \in \left(0, \frac{1-\lambda^U}{2-\lambda^U}\right)$$

and

$$y^S(t^*) < y^R(t^*) \text{ for } t^* \in \left(\frac{1-\lambda^U}{2-\lambda^U}, \frac{1}{2}\right) \text{ and no utilitarian has } t^* \text{ as a preferred tax rate}$$

Finally, from Proposition 4, given that the distribution of abilities $F(\omega)$ is positively skewed, we know that $t^* > 0$ and thus $y^S(t^*) < \overline{y}$, $y^U(t^*) < \overline{y}$, and $y^R(t^*) \lesseqgtr \overline{y}$. Therefore, we have that

1. If the poor get poorer,

$$\begin{cases} F(y) = G(y) \text{ for } y > y(\omega_1) \\ F(y) \leq G(y) \text{ for } y \leq y(\omega_1) \end{cases} \text{ and } y\left(\underline{\omega}\right) \text{ and } \overline{y} \text{ decrease}$$

   Hence, we have

   (a) If $y^S(t^*) \leq y(\omega_1)$, then $G\left(y^S(t^*)\right) \lesseqgtr F\left(y^S(t^*)\right)$, since $G(y) \geqslant F(y)$ for $y \leqslant y(\omega_1)$, and the effect of the decrease in $\overline{y}$ under $G(.)$ goes in the opposite direction (i.e. $\frac{\partial y^S(t^*)}{\partial \overline{y}} > 0$).
   If $y^S(t^*) > y(\omega_1)$, then $G\left(y^S(t^*)\right) < F\left(y^S(t^*)\right)$, since $\overline{y}$ is smaller under $G(.)$ than under $F(.)$.

   (b) If $y^R(t^*) \leqslant y(\omega_1)$, then $G\left(y^R(t^*)\right) \lesseqgtr F\left(y^R(t^*)\right)$, since $G(y) \geqslant F(y)$ for $y \leqslant y(\omega_1)$, and the decrease in $y(\underline{\omega})$ and in $\overline{y}$ under $G(.)$ go in opposite directions (i.e. $\frac{\partial y^R(t^*)}{\partial y(\underline{\omega})} < 0$ and $\frac{\partial y^R(t^*)}{\partial \overline{y}} > 0$).
   If $y^R(t^*) > y(\omega_1)$, then $G\left(y^R(t^*)\right) \lesseqgtr F\left(y^R(t^*)\right)$, since the decrease in $y(\underline{\omega})$ and in $\overline{y}$ under $G(.)$ go in opposite directions.

   (c) If $y^U(t^*) \leq y(\omega_1)$, then $G\left(y^U(t^*)\right) \lesseqgtr F\left(y^U(t^*)\right)$ (with strict equality if $t^* \geqslant \frac{1-\lambda^U}{2-\lambda^U}$), since $G(y) \geqslant F(y)$ for $y \leqslant y(\omega_1)$, and the effect of the decrease in $\overline{y}$ under $G(.)$ goes in the opposite direction (i.e. $\frac{\partial y^U(t^*)}{\partial \overline{y}} > 0$).
   If $y^U(t^*) > y(\omega_1)$, then $G\left(y^U(t^*)\right) \leqslant F\left(y^U(t^*)\right)$ (with strict equality if $t^* \geqslant \frac{1-\lambda^U}{2-\lambda^U}$), since $\overline{y}$ is smaller under $G(.)$ than under $F(.)$.

   Therefore, when the poor get poorer, the effect on the Condorcet winner tax rate t^* is ambiguous.

2. If the rich get richer,

$$\begin{cases} F(y) = G(y) \text{ for } y < y(\omega_2) \\ F(y) \geq G(y) \text{ for } y \geq y(\omega_2) \end{cases} \quad \text{and } y\left(\overline{\omega}\right) \text{ and } \overline{y} \text{ increase}$$

Hence, we have

(a) $G\left(y^S(t^*)\right) > F\left(y^S(t^*)\right)$, since $\overline{y}$ is higher under $G(.)$ than under $F(.)$.

(b) If $y^R(t^*) < y(\omega_2)$, then $G\left(y^R(t^*)\right) > F\left(y^R(t^*)\right)$, since $\overline{y}$ is higher under $G(.)$ than under $F(.)$.

If $y^R(t^*) \geqslant y(\omega_2)$, then $G\left(y^R(t^*)\right) \lesseqgtr F\left(y^R(t^*)\right)$, since $G(y) \leqslant F(y)$ for $y \geqslant y(\omega_2)$, and the effect of the increase in $\overline{y}$ under $G(.)$ goes in the opposite direction (i.e. $\frac{\partial y^R(t^*)}{\partial \overline{y}} > 0$).

(c) $G\left(y^U(t^*)\right) \geqslant F\left(y^U(t^*)\right)$ (with strict equality if $t^* \geqslant \frac{1-\lambda^U}{2-\lambda^U}$), since $\overline{y}$ is higher under $G(.)$ than under $F(.)$.

and thus

$$\alpha G\left(y^S(t^*)\right) + \beta G\left(y^R(t^*)\right) + \gamma G\left(y^U(t^*)\right) > \frac{1}{2} \text{ if } y^R(t^*) < y(\omega_2)$$

Therefore, $t^{**} > t^*$ whenever $y^R(t^*) < y(\omega_2)$.

3. If there is a mean-preserving spread,

$$\begin{cases} F(y) \leq G(y) \text{ for } y \leq y(\omega_1) \\ F(y) = G(y) \text{ for } y(\omega_1) \leq y \leq y(\omega_2) \\ F(y) \geq G(y) \text{ for } y \geq y(\omega_2) \end{cases} \quad \text{and } y(\underline{\omega}) \text{ decreases, } y(\overline{\omega}) \text{ increases, with } \overline{y} \text{ unchanged}$$

Hence, we have

(a) If $y^S(t^*) \leq y(\omega_1)$, then $G\left(y^S(t^*)\right) \geqslant F\left(y^S(t^*)\right)$, by definition of $F(.)$ and $G(.)$.

If $y^S(t^*) > y(\omega_1)$, then $G\left(y^S(t^*)\right) = F\left(y^S(t^*)\right)$, since $\overline{y}$ is the same under $G(.)$ as under $F(.)$.

(b) If $y^R(t^*) \leqslant y(\omega_1)$, then $G\left(y^R(t^*)\right) \geqslant F\left(y^R(t^*)\right)$, since $G(y) \geqslant F(y)$ for $y \leqslant y(\omega_1)$, and the effect of the decrease in $y(\underline{\omega})$ under $G(.)$ goes in the same direction (i.e. $\frac{\partial y^R(t^*)}{\partial y(\underline{\omega})} < 0$).

If $y(\omega_1) < y^R(t^*) < y(\omega_2)$, then $G\left(y^R(t^*)\right) > F\left(y^R(t^*)\right)$, since $y(\underline{\omega})$ is smaller under $G(.)$ than under $F(.)$.

If $y^R(t^*) \geqslant y(\omega_2)$, then $G\left(y^R(t^*)\right) \lesseqgtr F\left(y^R(t^*)\right)$, since $G(y) \leqslant F(y)$ for $y \geqslant y(\omega_2)$, and the effect of the decrease in $y(\underline{\omega})$ under $G(.)$ goes in the opposite direction.

(c) If $y^U(t^*) \leq y(\omega_1)$, then $G\left(y^U(t^*)\right) \geqslant F\left(y^U(t^*)\right)$ (with strict equality if $t^* \geqslant \frac{1-\lambda^U}{2-\lambda^U}$), by definition of $F(.)$ and $G(.)$.

If $y^U(t^*) > y(\omega_1)$, then $G\left(y^U(t^*)\right) = F\left(y^U(t^*)\right)$, since $\overline{y}$ is the same under $G(.)$ as under $F(.)$.

Hence, we have that

$$\alpha G\left(y^S(t^*)\right) + \beta G\left(y^R(t^*)\right) + \gamma G\left(y^U(t^*)\right) > \frac{1}{2} \text{ if } y^R(t^*) < y(\omega_2)$$

Therefore, $t^{**} > t^*$ whenever $y^R(t^*) < y(\omega_2)$.

□

Chapter 3

Interregional Transfers, Group Loyalty, and the Decentralization of Redistribution

3.1 Introduction

Separatist and/or decentralizing pressures are very often associated to both interregional income inequality and cultural heterogeneity between regions (see, for instance, Collier and Hoeffler (206)). Typically, there are two culturally homogeneous regions, and one of them is richer (such as Flanders in Belgium, or Catalonia in Spain). As a result of one region being richer, a centralized redistribution policy involves implicit interregional transfers taking place, and thus transfers of resources between individuals who do not share a common identity. The potential wish of the richer region to decentralize redistribution suggests two things: first, being richer, the region may want to implement its own redistribution policy which would be closer to the preferences of its population. Second, it may also be that the interregional transfers taking place through the centralized redistribution policy are not considered as legitimate by its population, and thus a decentralized system constitutes a way to get rid of it.

In turn, the presence of this two-sided heterogeneity between regions —average income and group identity— gives rise to the following trade-off regarding the choice between centralized and decentralized redistribution from a welfare perspective: on the one hand, a centralized system allows for a potentially beneficial pooling of national resources, thereby permitting to reduce interregional inequality. On the other hand, it also has a cost in the sense that individuals may be less willing to redistribute in a culturally divided society. That is, there may be a decrease in the degree of solidarity in the society under centralization.

In order to capture this trade-off, we set up a political economy model where individuals vote over a one-dimensional redistributive parameter, assuming that voters are utilitarian altruists and care relatively more about the well-being of individuals of their own region (i.e. there is group loyalty). We assume that there are two (culturally homogeneous) regions, and that one of them is richer in the sense that it has a higher proportion of rich individuals. As a result, under a centralized system of redistribution, there is a net transfer of resources from the rich to the poor region —this transfer being potentially undesirable from an individual point of view as a result of the lack of a common identity between the two regions.

The aim of the paper is both positive and normative. First, we characterize the equilibrium level of redistribution under both centralization and decentralization, and we show that while group loyalty always promotes redistribution under decentralization, it decreases the support for redistribution in a centralized system whenever the decisive voter is from the rich region.

Second, using the traditional welfaristic approach of fiscal federalism, we show that total welfare maximization is closely related to the minimization of inequality, both within and between regions. In particular, and unlike in the standard approach, the first-best solution is such that all individuals in the country end up with the same final consumption, no matter the strength of group loyalty nor the extent of interregional inequality.

Third, we analyze two particular cases under direct democracy, namely no interregional inequality and no group loyalty, in an attempt to isolate what we call the *scale effect* and the *pooling effect* of centralized redistribution, respectively. The scale effect relates to the fact that individuals are more willing to redistribute when redistribution is centralized, and are strictly better off by doing so in the absence interregional inequality. That is, independently of the existence of interregional transfers, centralizing redistribution has the additional benefit of increasing the willingness of rich voters to pay taxes, just because redistribution is implemented at a larger scale. This latter result is consistent with Oates' (1962) decentralization theorem (i.e. when there are regional spillovers and regions are identical), although it does not follow from the traditional internalization of spillovers argument.

In contrast, the pooling effect relates to the fact that individuals strictly prefer a centralized system of redistribution when they share a common group identity, since it allows them to pool national resources so as to homogenize redistribution in the country when there are income disparities between regions. This latter result contrasts sharply with the above-mentioned decentralization theorem, according to which the choice between centralization and decentralization implies a trade-off in that case (i.e. when there are regional spillovers and regions are not identical).

It turns out, then, that the rationale for decentralization in this political economy context only arises when both sources of regional heterogeneity are present (i.e. average income and group identity). Again, if the regions share a common identity (i.e. there is no group loyalty), centralization pareto-dominates decentralization, no matter whether the regions are identical or not. Likewise, whenever the two regions are equally rich, centralization pareto-dominates decentralization, even though individuals care relatively more about their own region. Using simulations of the model for the more general case of both interregional inequality and group loyalty, we show that the existence of the latter creates excessive pressures towards decentralization in the rich region from a welfare perspective. In particular, if the rich region can unilaterally decide to decentralize redistribution, and if group loyalty is strong enough, the political equilibrium is very likely to be inefficient. Furthermore, with full group loyalty, that is, in the absence of regional spillovers, it is not generally true that a decentralized system yields higher total welfare.

Finally, we show that, due to the presence of free-riding, allowing for voluntary transfers between regions under decentralization need not increase total welfare. Furthermore, simulations of the model show that centralization always welfare-dominates decentralization with transfers for all values of group loyalty for which such transfers occur.

There is a large body of experimental evidence showing that individuals tend to behave in an altruistic manner (see, for example, Fehr and Schmidt (1999) and Charness and Rabin (2002)). In this paper, we model preferences for redistribution based on utilitarian altruism, and, following Luttmer (2001), we assume that the strength of altruism towards specific individuals is determined by group loyalty. That is, individuals care mostly about the welfare of those belonging to their own group. As we assume a common identity within a region, this means that individuals care relatively more about both individuals and redistribution patterns within their own region.

There is a growing literature studying the effects of fractionalization along religious, ethnic or linguistic lines on public policy. In particular, the idea that support for redistribution may be lower in culturally diverse societies has been documented both empirically (see, among others, Alesina et al. (1999) and Luttmer (2001)) and theoretically (see, for instance, Austen-Smith and Wallerstein (2006) and Lind (2004)). Alesina et al. (1999) show that ethnic diversity tends to reduce both the supply of public goods and redistribution, and explain this fact based on heterogeneity of tastes. More generally, several theoretical reasons have been advanced in order to explain the detrimental effect of cultural diversity on public policy.

Here, cultural diversity is relevant to the extent that it might create high intra-group loyalty and less between-group loyalty. That is, we assume that individuals care relatively more —if not only— about the well-being of members of their own group (i.e. region). In that sense, our focus is not on the fact that *preferences* regarding public policy may differ across groups. A very related work is the one of Lind (2004), who highlights the fact that heterogeneity between individuals regarding both identity and income might have a joint impact on support for redistribution. In particular, he shows that while inequality within groups has the usual effect of promoting redistribution, inequality between groups has the opposite effect of reducing support for redistribution. However, as he does not assume that groups are geographically segmented, his focus is not on centralized versus decentralized redistribution.

The fact of regions being differently rich and the effect that this can have on decentralizing redistribution has already been investigated in a political economy context. For example, Persson and Tabellini (1994) show that if average income differs between regions, majority rule at the federal level produces less redistribution than at the local one. In their seminal paper on the breakup of nations, Bolton and Roland (1997) focus on the effects of regional heterogeneity regarding both average and median income on the incentives to secede, assuming that a breakup involves an efficiency cost. While in those papers, the authors assume regional heterogeneity regarding income distribution, they do not assume any type of cultural heterogeneity between regions nor altruistic motives for redistribution.

Now, the possibility that individuals exhibit group-loyal altruistic preferences, and the effect this has on the choice between centralized and decentralized redistribution, has also been investigated from a theoretical perspective. In his seminal work, Pauly (1973) shows that if redistribution is a spatially limited public good, it can be efficiently implemented at the local level. However, he does not address the issue of redistribution between regions, which clearly calls for some centralization. In Pauly's model, a decentralized policy has the advantage of being closer to regional tastes regarding redistribution. In that sense, his result can be seen as an application of Oates' decentralization theorem. In contrast, in our model, decentralization has the advantage of potentially increasing the degree of solidarity in the society, and hence redistribution. Another related work is the one of Bjorvatn and Cappelen (2006), where it is assumed that voters care about the poor only in their own community (i.e. there is full group loyalty), while decentralizing redistribution creates tax competition between jurisdictions, and hence a possible race-to-the-bottom. They show that the best level of government regarding redistribution is determined by a trade-off which depends on the nature of altruism (i.e. pure vs impure altruism). While Bjorvatn and Cappelen (2006) allow for tax competition under decentralization, they assume that jurisdictions are equally rich, and hence abstract from all issues related to interregional inequality.

Therefore, while the separate effects of the two sources of regional heterogeneity —average income and group identity— on the choice between a centralized and decentralized system of redistribution have already been analyzed theoretically, their joint impact has not yet been investigated. As we already argued earlier in this introduction, we believe that the *interactions* between regional differences regarding both average income and group identity have important implications for the issue of (de)centralizing redistribution, and this both from a positive and a normative point of view.

The rest of the paper is organized as follows: section 2 describes the setup. In section 3, we solve the model using the welfaristic approach of the traditional fiscal federalism literature. In section 4, we characterize the political equilibrium level of redistribution under both centralization and decentralization. Section 5 compares the welfare properties of centralized and decentralized redistribution under direct democracy. In section 6, we discuss our results in light of the traditional decentralization theorem in the context of redistribution. In section 7, we introduce the possibility of voluntary interregional transfers under decentralization. Section 8 concludes. All proofs can be found in the Appendix.

3.2 The Model

There are two regions A and B of equal size, and there are rich and poor individuals (voters) in both regions. All the poor individuals in the economy are endowed with income y^P, and all the rich with income y^R, where $y^R > y^P$. There are N individuals in total: n_A individuals in region

A, and n_B individuals in region B, where $n_A = n_B = n$. Furthermore, there are n_A^R rich and n_A^P poor individuals in region A, and n_B^R rich and n_B^P poor individuals in region B. Therefore, the total number of individuals in the economy is given by $N = n_A + n_B = 2n = \left(n_A^R + n_A^P\right) + \left(n_B^R + n_B^P\right)$. Only the rich individuals pay taxes, and only the poor individuals receive a transfer, on an egalitarian basis. The tax rate is linear and proportional, and there is no deadweight loss from taxation[1].

We introduce income inequality between regions by assuming that $n_A^P \neq n_B^P$ (and thus $n_A^R \neq n_B^R$), so that $\overline{y}_A = \frac{n_A^P y^P + n_A^R y^R}{n} \neq \overline{y}_B = \frac{n_B^P y^P + n_B^R y^R}{n}$, that is, average income differs between the two regions. If redistribution is decentralized, the regional tax rates t_A and t_B are decided regionally, and redistribution only occurs at the regional level (i.e. there are no interregional transfers). Conversely, if redistribution is centralized, there is only one tax rate t, which is decided nationally, and redistribution occurs at the national level, so that there are implicit transfers taking place from the rich to the poor region.

Notation 1. *From here on, $X(t_A, t_B)$ and $X(t)$ refer to X under decentralization and centralization respectively, while no brackets refer to X under any of the two institutional arrangements.*

The budget constraints of the rich and poor individuals in region $j = A, B$ under decentralization are thus given by

$$c_j^R(t_A, t_B) = (1 - t_j)\, y^R \tag{3.1}$$

$$c_j^P(t_A, t_B) = y^P + \frac{n_j^R t_j y^R}{n_j^P} \tag{3.2}$$

Under centralization, those constraints become

$$c^R(t) = (1 - t)\, y^R \tag{3.3}$$

$$c^P(t) = y^P + \frac{\left(n_A^R + n_B^R\right)}{\left(n_A^P + n_B^P\right)} y^R t \tag{3.4}$$

Individuals are utilitarian altruists, and altruism is determined by group loyalty. Formally, the utility function of, say, a rich individual in region A, is given by

$$U_A^R = u(c_A^R) + \alpha \left\{ \beta \left[\frac{n_A^P}{n} u(c_A^P) + \frac{n_A^R}{n} u(c_A^R) \right] + (1 - \beta) \left[\frac{n_B^P}{n} u(c_B^P) + \frac{n_B^R}{n} u(c_B^R) \right] \right\}$$

where $u(.)$ is strictly increasing and concave, $\alpha \in [0, 1]$ is a parameter which captures the intensity of individual altruistic motives (as opposed to selfish ones), and $\beta \in \left[\frac{1}{2}, 1\right]$ is the *group*

[1]The usual way to introduce an efficiency cost of taxation is to assume that the lump-sum transfer to individuals is given by $(t - \frac{t^2}{2})\overline{y}$, and this typically allows to avoid corner solutions for the derivation of the equilibrium tax rate under majority voting. Here, as we assume that private utilities are strictly concave, together with altruism, we are guaranteed to have an interior solution for the equilibrium tax rate even without this assumption. Furthermore, including such an efficiency cost would make the analysis intractable analytically. Even though we are aware that doing so would make the model more realistic, our focus here is entirely on distributional issues, and not on the distortions that taxation typically involves for the individual choice of labor supply.

loyalty parameter. Therefore, when $\beta = \frac{1}{2}$, an individual cares equally about welfare in both regions, whereas when $\beta = 1$, he only cares about welfare in his own region. We assume for tractability that private utility is logarithmic, that is, $u(c) = \ln c$. One verifies easily that preferences are single-peaked on the tax rate dimension under both centralization and decentralization, and thus the median voter theorem applies.

As already mentioned, average income heterogeneity between regions arises from the fact that $n_A^i \neq n_B^i$, $i = P, R$, so that there are interregional transfers taking place implicitly through a centralized system of redistribution. Furthermore, the (possible) lack of common identity between individuals across the two regions is captured by the group loyalty parameter β.

3.3 Inequality and Welfare: the First-Best Solution

In order to determine what would be an optimal solution in this economy under both centralization and decentralization, we first solve the model assuming that a benevolent social planner chooses the redistribution level under both systems so as to maximize total welfare. Assuming that poor individuals in both regions have the same utility function as the rich ones, total welfare is given by[2]

$$W = n_A^R U_A^R + n_A^P U_A^P + n_B^R U_B^R + n_B^P U_B^P$$

Simplifying this expression yields

$$W = (1+\alpha)\left[n_A^R u(c_A^R) + n_A^P u(c_A^P) + n_B^R u(c_B^R) + n_B^P u(c_B^P)\right]$$

Therefore, it turns out that total welfare is independent of group loyalty. Suppose that the social planner has to choose a uniform tax rate t so as to maximize $W(t)$, subject to the budget constraints (3.3) and (3.4). As total welfare is just the sum of private utilities, which are strictly concave, it turns out that the uniform tax rate that maximizes total welfare coincides with the one that minimizes total inequality (i.e. such that all individuals consume $\bar{c}$, where $\bar{c}$ is the average consumption in the country), and is given by

$$t^* = \frac{\left(y^R - y^P\right)}{y^R} \frac{\left(n_A^P + n_B^P\right)}{N}$$

Therefore, maximizing total welfare under a centralized system of redistribution means minimizing total inequality. Notice that this means minimizing inequality both within and between regions, that is, such that $c^R = c^P$ in both regions, and $\bar{c}_A = \bar{c}_B$. As we assume away the efficiency

[2]That is, the final utility of a poor individual in, say, region A, is given by

$$U_A^P = u(c_A^P) + \alpha \left\{ \beta \left[\frac{n_A^P}{n} u(c_A^P) + \frac{n_A^R}{n} u(c_A^R) \right] + (1-\beta) \left[\frac{n_B^P}{n} u(c_B^P) + \frac{n_B^R}{n} u(c_B^R) \right] \right\}$$

loss arising from taxation, and as private utilities are strictly concave, total welfare is maximized when consumption of rich and poor individuals is equalized both within and between regions[3]. Observe, furthermore, that this is true no matter the strength of group loyalty. That is, even though voters in both regions care mostly about welfare in their own region, the first-best solution under centralization is always perfect equality[4].

Suppose now that the same social planner has to choose (t_A, t_B) so as to maximize $W(t_A, t_B)$ subject to the budget constraints (3.1) and (3.2). Observe that in this case, there are no transfers of resources between regions. The optimal tax rate in region $j = A, B$ is then given by

$$t_j^* = \frac{(y^R - y^P)}{y^R} \frac{n_j^P}{n}$$

and is such that inequality is minimized within the region. That is, t_j^* is such that all individuals in region j consume $\bar{c}_j$. Notice, however, that inequality remains between regions since $\bar{c}_A \neq \bar{c}_B$.

Comparing the two solutions, we then get the following result:

Proposition 1 (the First-Best Solution). *Assume $n_i^P < n_j^P$, $i, j = A, B$. In that case, $t_i^* < t^* < t_j^*$, and total welfare is strictly higher under the centralized solution for all β.*

Therefore, the optimal centralized solution welfare-dominates the decentralized one. The above result is somewhat striking, as it basically states that when the two regions are not identical, even though the regional spillovers can be fully internalized, the uniform solution yields higher total welfare. This contrasts with the traditional result of the fiscal federalism literature, according to which decentralization should be better in that case. In fact, it is precisely because $n_A^P \neq n_B^P$ that centralization is more efficient, since it allows to eliminate inequality between regions. Indeed, notice that if $n_A^P = n_B^P$, it follows that $t_A^* = t_B^* = t^*$ and $W(t^*) = W(t_A^*, t_B^*)$. In other words, when the only source of heterogeneity between regions is the fact that $n_A^P \neq n_B^P$, the uniformity under centralization does not constitute a cost —to the contrary. As a result, centralizing redistribution is unambiguously —and always— better, independently of the strength of group loyalty.

[3]Including an efficiency cost of taxation would clearly matter for total welfare, as total consumption would then be decreasing in t. A conjecture is that choosing the optimal t^* would then imply solving a trade-off between efficiency and inequality. In particular, t^* would be such that $c^R(t^*) > c^P(t^*)$ and thus $\bar{c}_A \neq \bar{c}_B$. As mentioned earlier, our focus here is on distributional issues rather than on the inefficiencies arising from taxation.

[4]Observe, however, that this is the case because the only source of heterogeneity between regions is the fact that $n_A^P \neq n_B^P$. Assuming that $\beta_A \neq \beta_B$, total welfare is then given by

$$W = (1+\alpha)\left[n_A^R u(c_A^R) + n_B^R u(c_B^R) + n_A^P u(c_A^P) + n_B^P u(c_B^P)\right] + (\beta_A - \beta_B)\left[n_A^R u(c_A^R) - n_B^R u(c_B^R) + n_A^P u(c_A^P) - n_B^P u(c_B^P)\right]$$

Alternatively, assuming that $\alpha_A \neq \alpha_B$, total welfare becomes

$$W = \left[n_A^P u(c_A^P) + n_A^R u(c_A^R)\right][1 + \alpha_B + \beta(\alpha_A - \alpha_B)] + \left[n_B^P u(c_B^P) + n_B^R u(c_B^R)\right][1 + \alpha_A - \beta(\alpha_A - \alpha_B)]$$

In both cases, the optimal solution would depend on the group loyalty parameter β, and would involve some strictly positive level of (total) inequality.

Observe, finally, that neither the first-best solution t^* nor the second-best one (t^*_A, t^*_B) will be attained in a direct democracy, the reason being that voters are at least partly self-interested.

In order to compare the relative benefits of centralized and decentralized redistribution under direct democracy, we now characterize the political equilibrium level of redistribution of this economy for both institutional arrangements.

3.4 Equilibrium Redistribution under Direct Democracy

In order to focus on altruistic motives for redistribution and donor motivation, we assume that the decisive voter is always a rich individual under both centralization and decentralization. As the poor do not pay taxes, redistribution here is meant to be interpreted as assistance to the poor by the rich. In that sense, our focus is not on whether the poor are able to expropriate the rich through their political power, but rather on the incentives of the rich to decentralize redistribution to the poor given the presence of interregional inequality and group loyalty[5].

3.4.1 Political Equilibrium under Decentralization

In order to choose the regional equilibrium tax rate, the decisive (rich) voter in region $j = A, B$ maximizes $U_j^R(t_A, t_B)$ with respect to t_j. The corresponding first-order condition is given by

$$u'\left(c_j^R\right) = \alpha\beta \frac{n_j^R}{n}\left[u'\left(c_j^P\right) - u'\left(c_j^R\right)\right]$$

For $u(c) = \ln c$, the equilibrium tax rate in region $j = A, B$ under decentralization is given by

$$t_j = \frac{\alpha\beta}{(1+\alpha\beta)}\frac{n_j^P}{n}\frac{\left(y^R - y^P\right)}{y^R} - \frac{1}{(1+\alpha\beta)}\frac{n_j^P}{n_j^R}\frac{y^P}{y^R}$$

and it has the following properties[6]:

- $\frac{\partial t_j}{\partial n_j^P} > 0$ if and only if $\alpha\beta > \frac{y^P}{(y^R - y^P)}\left(\frac{n}{n_j^R}\right)^2$
- $\frac{\partial t_j}{\partial (y^R - y^P)} > 0$
- $\frac{\partial t_j}{\partial \alpha} > 0$
- $\frac{\partial t_j}{\partial \beta} > 0$

[5]Given that individuals exhibit altruistic preferences, observe that we do not need the poor to be decisive in order to have positive redistribution in equilibrium. Notice, furthermore, that if the poor were decisive —and given that there is no efficiency loss from taxation— the equilibrium level of redistribution would be such that the poor end up consuming strictly more than the rich, which is unrealistic.

[6]See section 1 of the Appendix for the derivatives computations.

The equilibrium tax rate is increasing in the number of poor if and only if altruism and group loyalty are high enough[7]. In particular, the higher pre-tax income dispersion and/or the proportion of rich, or, in other words, the cheaper it is to redistribute, the more likely that the condition is satisfied. The tax rate is increasing in pre-tax income dispersion, that is, as in the standard model of voting on redistribution (Meltzer and Richards (1981)), more pre-tax income inequality leads to more redistribution in equilibrium[8]. Finally, when α increases, the median voter becomes relatively more altruistic, which increases his preferred tax rate. Observe that an increase in the group loyalty parameter β has exactly the same effect as an increase in α. Indeed, as there is no possibility of interregional transfers, the group loyalty parameter does not differ in essence from the altruistic weight α under decentralization. Therefore, stronger group loyalty is equivalent to stronger altruism in this case, meaning a higher willingness to redistribute in equilibrium.

3.4.2 Political Equilibrium under Centralization

Assume, without loss of generality, that the decisive voter under centralization is from region A. Provided that the poor individuals do not vote on the transfers to themselves, if $n_A = n_B$ and $n_A^R > n_B^R$, it follows directly that the decisive voter under centralization is from region A (i.e. the rich region). However, in order to be more general, we would like to allow for the possibility of the poor region being decisive under centralization as well. This could be the case, for instance, if we allow population size to differ across regions. However, as we do not want a particular region to have more weight when evaluating aggregate welfare, we will keep the assumption of equal region size, and just assume that either the poor or the rich region can be decisive under centralization[9].

As in the decentralized case, the decisive (rich) voter (from region A) implements his preferred tax rate so as to maximize $U_A^R(t)$. The corresponding first-order condition is given by

$$\frac{\left(n_A^R + n_B^R\right)}{\left(n_A^P + n_B^P\right)}\alpha\left[\beta\frac{n_A^P}{n}u'\left(c_A^P\right) + (1-\beta)\frac{n_B^P}{n}u'\left(c_B^P\right)\right] = u'\left(c_A^R\right) + \alpha\left[\beta\frac{n_A^R}{n}u'\left(c_A^R\right) + (1-\beta)\frac{n_B^R}{n}u'\left(c_B^R\right)\right]$$

Assuming that $u(c) = \ln c$, the equilibrium tax rate is given by

$$t = \frac{1}{(1+\alpha)}\left\{\alpha\left[\beta\frac{n_A^P}{n} + (1-\beta)\frac{n_B^P}{n}\right] - \frac{y^P}{y^R}\frac{\left(n_A^P + n_B^P\right)}{\left(n_A^R + n_B^R\right)}\left[1 + \alpha\beta\frac{n_A^R}{n} + \alpha(1-\beta)\frac{n_B^R}{n}\right]\right\}$$

[7] Notice that when taking the derivative with respect to the number of poor individuals, we keep region size constant. That is, we look at the effect of an increase in the regional *proportion* of poor.

[8] Observe, however, that the positive relationship between inequality and the level of redistribution here occurs as a result of inequality aversion (since private utilities are strictly concave), while it is due to selfish motives in the traditional model with self-interested voters (since more inequality means that the median voter is poorer, and hence benefits more from redistribution).

[9] Note that the two regions having different sizes does not have any effect on voting decisions, as it is the regional *proportions* of rich and poor that enter individual preferences.

and it has the following properties[10]:

- $\frac{\partial t}{\partial n_A^P} > 0$ if and only if $\beta > \beta_1$
- $\frac{\partial t}{\partial n_B^P} > 0$ if and only if $\beta < \beta_2 < 1$
- $\frac{\partial t}{\partial (y^R - y^P)} > 0$
- $\frac{\partial t}{\partial \alpha} > 0$
- $\frac{\partial t}{\partial \beta} > 0$ if and only if $n_A^P > n_B^P$

The centralized tax rate is increasing in the proportion of local poor if β is strictly higher than some threshold, while it is increasing in the proportion of poor in B if β is strictly lower than some (other) threshold. Observe that when $\beta = 1$, $\frac{\partial t}{\partial n_B^P} < 0$ for any other parameter values. That is, with full group loyalty, and given that region A is decisive, the centralized tax rate is always decreasing in the proportion of poor in region B. Finally, as in the decentralized case, the tax rate is increasing in both pre-tax income dispersion and general altruism.

More importantly, we have the following result regarding the effect of group loyalty on the equilibrium tax rate when redistribution is centralized:

Proposition 2. *Assume region A is decisive under centralization. Then, the equilibrium level of redistribution is decreasing in the strength of group loyalty if and only if A is the rich region.*

If region A is the rich (and decisive) region in the sense that the proportion of rich individuals is higher than in B, an increase in group loyalty decreases the equilibrium tax rate. As the decisive voter cares relatively more about his own region, the fact that most (i.e. more than 50%) of the global tax revenue goes to the other region induces him to implement a lower tax rate in equilibrium. Said in other words, the stronger group loyalty, the more the rich in the rich region are willing to sacrifice redistribution between individuals in order to decrease redistribution between regions.

Conversely, if region A is the poor region (i.e. $n_A^P > n_B^P$), an increase in β increases the equilibrium tax rate, as the decisive voter wants to take advantage of the implicit transfers from the rich region to a greater extent. Said in other words, the stronger group loyalty, the more the rich in the poor region are willing to sacrifice their own income in order to increase redistribution between regions.

3.5 Welfare under Direct Democracy

We now turn to the analysis of the relative benefits of centralized and decentralized redistribution under direct democracy. To do so, we first focus on two particular cases, namely the ones involving only one type of heterogeneity between regions. More specifically, we compare centralization and

[10]See section 1 of the Appendix for the derivatives computations.

decentralization in the absence of interregional income inequality, and in the absence of group loyalty, respectively. Doing so allows us to further understand how income inequality between regions on the one hand, and group loyalty on the other hand, interact to shape the level of redistribution under both institutional arrangements, and which of the two yields higher total welfare. As we show below, it turns out that when only one type of heterogeneity is present, centralization always Pareto-dominates decentralization.

3.5.1 No Interregional Inequality: the Scale Effect of Centralization

Assume that all parameters are the same across the two regions, and in particular that $n_A^P = n_B^P = n^P$ (and thus $n_A^R = n_B^R = n^R$). Under decentralization, the regional equilibrium tax rates are identical, and are given by

$$t_A = \frac{\alpha\beta}{(1+\alpha\beta)} \frac{n^P}{n} \frac{\left(y^R - y^P\right)}{y^R} - \frac{1}{(1+\alpha\beta)} \frac{n^P}{n^R} \frac{y^P}{y^R} = t_B$$

Under centralization, no matter which region the median voter belongs to, the uniform tax rate is given by

$$t = \frac{\alpha}{(1+\alpha)} \frac{n^P}{n} \frac{\left(y^R - y^P\right)}{y^R} - \frac{1}{(1+\alpha)} \frac{n^P}{n^R} \frac{y^P}{y^R}$$

As the burden of redistribution is shared equally among the rich individuals in the two regions, the centralized tax rate is now independent of β. Hence, the presence of group loyalty *per se* cannot explain why redistribution would be lower in culturally diverse societies, a point already made by Bjorvatn and Cappelen (2006). In addition to group loyalty, some heterogeneity between the regions is needed regarding how rich they are, implying in turn that implicit interregional transfers take place as a result of redistribution being centralized.

Proposition 3 (the Scale Effect of Centralized Redistribution)**.** *If $n_A^P = n_B^P$, it follows that $t_A = t_B < t$, unless $\beta = 1$, in which case $t_A = t_B = t$. If $\beta < 1$, then, all individuals in the economy are strictly better off under centralized redistribution (i.e. centralization Pareto-dominates decentralization).*

When the two regions are equally rich, and as long as individuals care about welfare in the other region, the centralized tax rate is strictly higher than the regional ones.

Why is this so? As no interregional transfers are allowed under decentralized redistribution, even though the median voter in each region may want to transfer part of the regional tax revenue to the poor in the other region, he cannot. One may then believe that the latter constraint of the decentralized system makes the regional median voters prefer a strictly lower tax rate than under centralization. Observe, however, that this will be the case only when there are income disparities between regions. Indeed, whenever the two regions are equally rich, centralizing redistribution does not involve any (net) transfer of resources between the two regions. Furthermore, the very rationale for voluntary interregional transfers, clearly, only arises if there is a poor and a rich region[11]. Therefore, there seems to be some other mechanism at work leading to lower redistribution under decentralization in the absence of interregional income inequality.

Under centralization, the decisive voter can force every other rich in the country to pay his preferred tax rate, the revenue of which is shared equally among the poor everywhere in the country. Under decentralization, each median voter can only force the rich within the region to pay the tax, the revenue of which is transferred to the local poor. Thus, under the centralized regime, and given that $n_A^P = n_B^P$, the decisive voter makes twice as many rich redistribute to twice as many poor. As a result, given that redistribution is implemented at this larger scale, the decisive voter (as any other rich) is willing to redistribute strictly more when redistribution is centralized —which in turn is beneficial to everyone. This is what we call the *scale effect* of centralized redistribution.

Observe that here, the mechanism through which centralization yields a better outcome — in the sense of more redistribution (which is profitable for everyone)— is very different from the one at work in the traditional normative approach described by the fiscal federalism literature. Traditionally, the benefit of centralization lies in the fact that, under such a system, the central government —whose objectives are welfaristic— is able to internalize the regional spillover effects of public good provision. Here, in contrast, the voter choosing the tax rate under both centralization and decentralization is the same individual, who "selfishly" maximizes his own utility function when doing so. Therefore, our decisive voter, when choosing the centralized tax rate, is not internalizing anything. In particular, he does not internalize the fact that individuals in the other region also value welfare in his own region to some extent (i.e. the regional spillovers of redistribution). Again, the reason why the median voter chooses to redistribute more under centralization (as long as $\beta < 1$) lies in the fact that, under such a system, he can force more rich to redistribute to more poor (the scale effect of centralization). Said in other words, the under-provision of redistribution under decentralization here comes from the reduced scale at which redistribution is implemented from the point of view of the decisive voter, while it comes from the non-internalization of spillovers in the standard normative approach of fiscal federalism.

3.5.2 No Group Loyalty: the Pooling Effect of Centralization

Assume now that $\beta = 1/2$, that is, there is no group loyalty. In this case, it turns out that the final utility of any voter $i = R, P$ in region $j = A, B$ is an increasing function of total welfare under the corresponding solution:

$$U_j^i = u(c_j^i) + \frac{\alpha}{(1+\alpha)} \frac{W}{N}$$

Therefore, when $\beta = 1/2$, each individual basically cares about two things: himself and total welfare. Observe that, since poor and rich individuals now have the same final utility across regions,

[11] We allow for the possibility of voluntary interregional transfers under decentralized redistribution in section 7.

the preferred centralized tax rate of a rich voter in A and in B is the same, even though $n_A^P \neq n_B^P$. This is true only when $\beta = 1/2$, since in that case, the potential disutility of implicit interregional transfers disappears in the rich region, while individuals in the poor region do not wish to exploit the rich one any longer. As a result, individuals in the rich and the poor region now agree on the optimal level of redistribution to implement under a centralized system, both within and between regions.

Given that each individual weights equally welfare in the two regions, the decisive voter, independently of the region he belongs to, would like to pool national resources and choose the tax rate so as to implement his preferred redistribution policy everywhere in the country. Indeed, as $\beta = 1/2$, the decisive voter does not wish to discriminate the other individuals according to the region they belong to. The fact that all the rich in the nation would choose the same centralized tax rate confirms this intuition: given the absence of group loyalty, the fact that $n_A^P \neq n_B^P$ is not relevant any longer for the choice of t, as the only thing that matters from the point of view of any individual is the *national* proportion of poor, and, as a result, the median voter would like to smooth their consumption across the nation. In fact, observe that voters now partly act as a social planner, to an extent depending on the weight they attach to selfish versus altruistic motives.

A decentralized system, therefore, just adds a constraint to the choice of the decisive voter when $\beta = 1/2$, as he can only choose redistribution in his region, even though he cares equally about the other region. Of course, this means that the scale effect of centralization, which we were able to isolate in the previous section by assuming no interregional inequality, is also at work here. However, the fact that now $n_A^P \neq n_B^P$ makes the ordering of the regional and national tax rates ambiguous, so that it is not clear whether the median voter is always willing to increase the tax rate as a result of redistribution being implemented at a higher scale[12].

More importantly, what we want to stress here —for the case of $\beta = 1/2$ together with $n_A^P \neq n_B^P$— is the fact that, besides the scale of redistribution being reduced under decentralization, such a system has the additional constraint of not allowing transfers of resources between regions. As individuals no longer give priority to their own region, they consequently attach the same weight to reducing inequality within and between the two regions. Therefore, the additional interregional inequality which arises as a result of decentralization (i.e. the fact that both $c_A^R(t_A, t_B) \neq c_B^R(t_A, t_B)$ and $c_A^P(t_A, t_B) \neq c_B^P(t_A, t_B)$) constitutes a pure loss from the point of view of any individual, and thus from an aggregate welfare perspective. This is what we call the *pooling effect* of centralized redistribution.

Proposition 4 (the Pooling Effect of Centralized Redistribution)**.** *Assume region A is decisive under centralization and $n_A^P \neq n_B^P$. In the absence of group loyalty (i.e. $\beta = 1/2$), all individuals in the economy are strictly better off under centralized redistribution (i.e. centralization Pareto-dominates decentralization).*

[12]Observe, however, that for $\beta = 1/2$, the average consumption of the poor is strictly higher under centralized redistribution (i.e. $c^P(t) > \left[c_A^P(t_A, t_B) + c_B^P(t_A, t_B)\right]/2$). Therefore, although it is not clear whether the rich are willing to pay more taxes under centralization, the absolute level of redistribution is indeed higher.

In the next section, we turn to the general case of both interregional inequality (i.e. $n_A^P \neq n_B^P$) and group loyalty (i.e. $\beta > 1/2$). Given the complexity of the algebra involved, we could not derive analytical results for this more general case. However, by simulating the model for different parameter values, we are able to give some intuition regarding the welfare implications of allowing both types of regional heterogeneity —average income and group identity— to interact.

3.5.3 Interregional Inequality and Group Loyalty: Simulations

In the two previous sections, we analyzed two particular cases —no interregional inequality and no group loyalty— in an attempt to isolate what we call the scale effect and the pooling effect of centralized redistribution, respectively. While the scale effect relates to the fact that individuals are willing to redistribute more when redistribution is centralized, and are strictly better off by doing so in the absence interregional inequality, the pooling effect relates to the fact that individuals strictly prefer a centralized system of redistribution when they share a common group identity, since it allows them to pool national resources so as to homogenize redistribution in the country.

In this section, we turn to the more general case of both interregional income inequality (i.e. $n_A^P \neq n_B^P$) and group loyalty (i.e. $\beta > 1/2$). Clearly, both the scale and the pooling effect of centralization are at work here as well. However, they now interact in a way that renders things somewhat more complex.

As before, the stronger group loyalty, the smaller the scale effect of centralization, since individuals care relatively less about general welfare (and thus redistribution) in the other region. Then, the stronger group loyalty, the smaller the pooling effect of centralization in the rich region, and the bigger in the poor one. Indeed, the stronger group loyalty, the less the rich region values interregional transfers, whereas the more the poor region values the latter transfers (although for selfish reasons).

When $\beta > 1/2$ and $n_A^P \neq n_B^P$, the rich individuals in the two regions do not agree any longer on the best centralized tax rate to implement, and in fact, the higher β, the more their preferences diverge (recall that $\partial t/\partial\beta > 0$ if and only if $n_A^P > n_B^P$). More specifically, as group loyalty gets stronger, the poor region is increasingly willing to exploit the rich region through implicit interregional transfers, so that when it is decisive, (1) redistribution increases, and (2) it is better off under centralization. Conversely, as group loyalty gets stronger, the rich region is increasingly unwilling to implement the latter transfers, so that when it is decisive (1) redistribution decreases, and (2) it is more and more likely to prefer decentralization[13].

As before, let assume that the decisive voter under centralization is from region A, and $\beta > 1/2$. If $n_A^P < n_B^P$, the median voter now faces a trade-off when choosing the tax rate under the centralized regime. On the one hand, he can force the whole population of rich to pay a given tax rate in order to help the poor, which makes him more willing to redistribute (the scale effect). On the other

[13] In the limit, when there is full group loyalty (i.e. $\beta = 1$) and $n_A^P > n_B^P$, all voters in region A are strictly better off under centralization, whereas if $n_A^P < n_B^P$, all voters in region A are strictly better off under decentralization.

hand, he knows that most contributions will go to the region about the one he cares the least, which makes him less willing to redistribute (the pooling effect). In contrast, if the poor region is decisive under centralization (i.e. $n_A^P > n_B^P$), the decisive voter faces no such a trade-off, since both the scale and the pooling effects go in the sense of more redistribution.

The properties of total welfare regarding group loyalty under both centralization and decentralization are given in the following proposition:

Proposition 5 (Group Loyalty and Total Welfare)**.** *Total welfare under decentralization is increasing in group loyalty. Assume region A is decisive under centralization. Then, total welfare under centralization is increasing in group loyalty if and only if A is the poor region.*

Under decentralization, group loyalty is clearly valuable for society, as it increases solidarity within regions (the scale effect is decreasing in both regions). Likewise, it is beneficial under centralization when the poor region is decisive, as it increases solidarity both within and between regions (the pooling effect is increasing in the poor (and decisive) region). Finally, group loyalty is harmful under centralization when the rich region is decisive, as it decreases solidarity both within and between regions (the pooling effect is decreasing in the rich (and decisive) region).

When the poor region is decisive under centralization (i.e. $n_A^P > n_B^P$), simulations of the model show that total welfare is strictly higher under centralized redistribution for all β. Indeed, although $W(t)$ and $W(t_A, t_B)$ are both increasing in β in that case, the former increases faster as the rise in t allows interregional inequality to decrease accordingly. Conversely, if $n_A^P < n_B^P$, it follows that $W(t)$ is decreasing in β, so that centralization welfare-dominates decentralization whenever β is smaller than some threshold $\beta_3 \lesseqgtr 1$. If $\beta_3 > 1$, centralization always yields higher welfare than decentralization, no matter which region is decisive under centralization.

Figure 3.1 depicts $W(t) - W(t_A, t_B)$ as a function of β and n_B^P, for $n_A^P = 10$ (and thus $n_A^P < n_B^P \in [20, 90]$). Observe that for all n_B^P, $W(t) - W(t_A, t_B)$ is decreasing in β. Furthermore, $\beta_3 = 0.75$ when $n_B^P = 20$, while $\beta_3 = 0.96$ when $n_B^P = 90$. That is, a bigger proportion of poor in the poor region makes it more likely that centralization welfare-dominates decentralization even for very strong group loyalty, since the benefit of reducing inequality between regions more than offsets the cost of increasing inequality between individuals (t small).

After having investigated which institutional arrangement yields higher total welfare, a natural question that arises is whether the best solution is sustainable as an equilibrium in a direct democracy. In turn, this requires making additional assumptions regarding the rule under which a given system —centralized or decentralized— can be implemented. Suppose centralization is the status quo. Implementing a decentralized system could require, for instance, the following: unanimity in both regions, majority in both regions, majority in one region, majority in the country. Then, if the best solution is not sustainable as an equilibrium under the corresponding rule, another question that arises is whether there exists an accommodating policy (i.e. a tax rate) such that the inefficient solution (i.e. the one that is welfare-dominated) can be avoided.

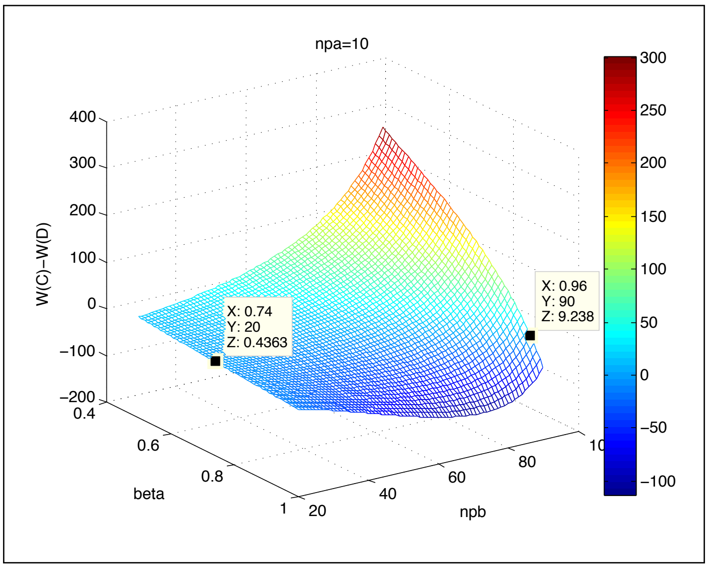

Figure 3.1: Total welfare under centralization and decentralization

When the decisive voter under centralization is from the rich region, his relative well-being under the two alternative systems depends on the following trade-off: decentralization makes reducing inequality within his region cheaper, but does not allow for interregional transfers. Conversely, centralization allows both to reduce interregional inequality and increase the scale of redistribution, but makes reducing inequality within his region more costly. Therefore, whether the decisive (rich) voter in the rich region is better off under a decentralized system depends on the resolution of this trade-off, and hence on β. In particular, the stronger group loyalty, the more likely that the decisive voter is better off under decentralized redistribution.

In Figure 3.2, we represented $U_A^i(t) - U_A^i(t_A, t_B)$, $i = R, P$, together with $W(t) - W(t_A, t_B)$ when the rich region is decisive under centralization (i.e. $n_A^P < n_B^P$). Observe that the higher β, the more likely that both the poor and rich individuals in the rich region strictly prefer decentralization to centralization. Likewise, the higher β, the more likely that the latter is welfare-dominated. However, there is some range of β for which total welfare is strictly higher when redistribution is centralized, but for which decentralization obtains in equilibrium, provided that the rich region can unilaterally decide to decentralize. Furthermore, notice that for this same range of β, there obviously never exists an accommodating tax rate such that decentralization can be avoided.

Therefore, when the rich region is decisive, an inefficiency arises: for some range of β, decentralization obtains in equilibrium although it is welfare-dominated.

When the poor region is decisive and group loyalty is strong, even though $W(t) > W(t_A, t_B)$, one may wonder whether it is reasonable to promote a centralized system of redistribution. Indeed, the higher welfare arising under centralization here comes from the poor region selfishly exploiting the rich one, which should not be considered as something desirable. If a majority in each region is required to implement a decentralized system, and if the rich region is better off under decen-

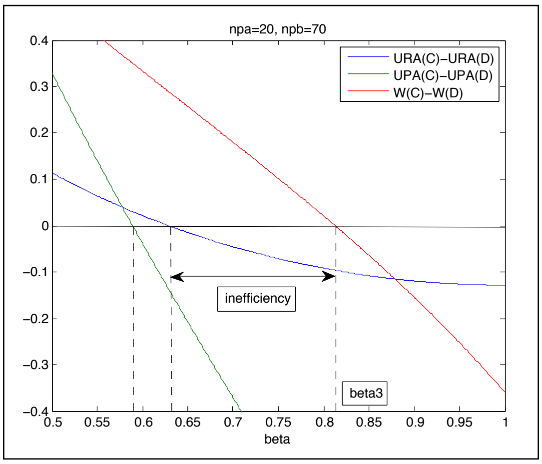

Figure 3.2: Group loyalty and welfare when the rich region is decisive

tralization, it will then never be able to stop making implicit transfers to the poor region through the centralized redistribution policy. In turn, such a situation may be rather unstable and costly, as separatist tensions are very likely to arise.

However, if group loyalty is not "too" strong, it turns out that there may exist an accommodating (centralized) tax rate $\widetilde{t} < t$ such that the rich in both regions are better off under $\widetilde{t}$ than under decentralization, and such that total welfare remains higher.

In order to illustrate this fact, let us look at an example. Suppose that $n_A^P = 90$ and $n_B^P = 10$. Suppose, furthermore, that $\beta = 0.73$ (i.e. there is imperfect group loyalty). Figure 3.3 depicts individual utility levels of the rich in both regions under decentralization and centralization for $t \in [0, 0.8]$. In this case, $W(t) > W(t_A, t_B)$ as long as $t > 0.045$. At the equilibrium t, we have $W(t) > W(t_A, t_B)$, but $U_B^R(t_A, t_B) > U_B^R(t)$, so that decentralization obtains in equilibrium if the rich in B can unilaterally decide to decentralize. However, as can be seen in the figure, there exists $\widetilde{t} < t$ such that both $U_j^R(\widetilde{t}) > U_j^R(t_A, t_B)$, $j = A, B$, and $W(\widetilde{t}) > W(t_A, t_B)$, so that decentralization —the inefficient solution— can be avoided. Notice that a necessary condition for an accommodating tax rate to exist is that the rich in B would be better off under centralization if he was decisive (i.e. β cannot be too high).

As a conclusion of this section, it appears that the existence of group loyalty creates excessive pressures towards decentralization in the rich region from a welfare perspective. In particular, if the rich region can unilaterally decide to decentralize redistribution, and if group loyalty is strong enough, the political equilibrium is very likely to be inefficient.

3.6 Redistribution, Group Loyalty and the Decentralization Theorem

One of the traditional arguments against centralization is that such a system is less sensitive to regional preferences —the so-called *preference-matching* argument. Typically, such an inefficiency is

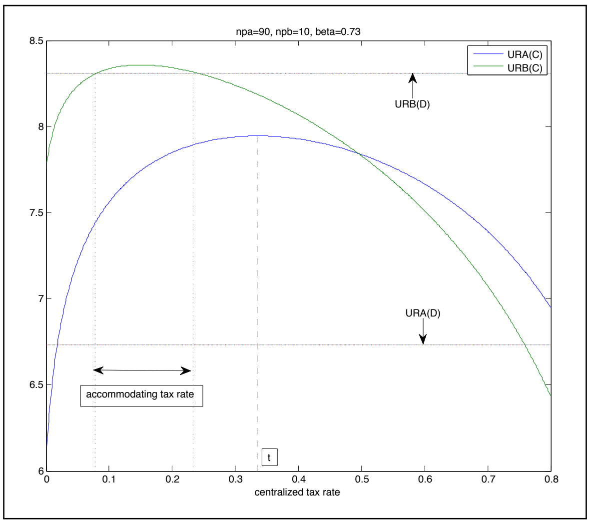

Figure 3.3: Existence of an accommodating tax rate when the poor region is decisive

generated by the "uniformity" assumption under centralization. However, this *ad hoc* assumption is not necessary to generate reduced preference-matching under centralization. For example, Lockwood (2002) highlights a similar inefficiency in a political economy model where the full political process is modeled and where no uniformity is assumed under centralization. While this uniformity assumption has been much criticized both on empirical and theoretical grounds (see, for instance, Besley and Coate (2003)), we believe that it remains appropriate in our set up. Indeed, as here the purpose of policy is pure redistribution, it is quite natural to assume a rule of horizontal equity within the geographical area in which the policy is implemented. That is, individuals should be treated the same way under centralization independently of their geographical location.

How does, then, the preference-matching argument translate in our set up? In the fiscal federalism literature, the cost of centralization typically arises because the regions value public goods differently. In particular, if there are two regions, one of them values the public good more than the other, and the central government —whose objectives are welfaristic— when setting the uniform level under centralization, consequently under-provides or over-provides the public good in a given region. In our setup, all voters value redistribution the same way (i.e. α is the same for all voters, both within and between regions) and the only source of heterogeneity is the fact that one region has a higher proportion of poor individuals than the other. As a result of this heterogeneity, the equilibrium regional tax rates indeed differ under decentralization, that is, $t_A \neq t_B$ as long as $n_A^P \neq n_B^P$.

Bolton and Roland (1997) (henceforth BR) have shown that income-based redistribution has two effects on the incentives to secede: a political effect, as the regional and national median incomes differ, and a tax base effect, as average income differs between regions. The political effect reflects differences in preferences for redistribution, and always induces a given region to secede, independently of the existence of interregional transfers. Such transfers arise when regional average incomes differ, and typically induce richer regions to secede (the tax-base effect).

The political effect relates to the preference-matching argument described above. In BR, median incomes differ across regions, and thus $t_A \neq t_B$. In our setup, the proportion of poor individuals differ across regions, and thus $t_A \neq t_B$ as well. Therefore, in both cases, regions are heterogeneous, which translates into different preferences regarding redistribution under decentralization. However, while the political effect always induces any region to secede in BR, things get much more complex when one allows for the possibility of altruism and group loyalty.

The crucial difference between our setup and the one of BR lies in the fact that, with altruism, the regional median voters have different preferences only when there is group loyalty. If this is not the case ($\beta = 1/2$), it still holds that $t_A \neq t_B$ as long as $n_A^P \neq n_B^P$, but this is only due to the fact that the median voters are not able to pool national resources, as they would otherwise vote for the same tax rate under a centralized system, and would be strictly better off by doing so. That is, the two median voters do not differ in their *tastes*, but rather in their *ability* to redistribute. Said in other words, the presence of altruism in our model sort of cancels out the regional differences in redistributive preferences as long as all individuals in the country share a common identity.

In the presence of group loyalty (i.e. $\beta > 1/2$), however, redistributive preferences under both centralization and decentralization now differ between regions, and both regions may strictly prefer a decentralized system. The rich region, because it dislikes —to some extent— the interregional transfers taking place under centralization. And the poor region, because the centralized tax rate might be so low (if the rich region is decisive) that it may actually prefer to implement its own redistribution policy even though it does not benefit from a transfer. In any case, notice that it is the *interaction* between group loyalty and regional differences in average income that generate the incentives to decentralize[14].

Finally, in BR, the tax-base effect always induces the richer region to secede. Again, this is true in our model only to the extent that the median voter in the richer region cares more (or only) about his own region. Otherwise, if there is no group loyalty, the pooling of national resources is beneficial to everyone. As in BR, the tax-base effect always reduces the poor region's incentives to decentralize in our model. However, when the rich region is decisive and group loyalty is very strong, the poor region may well be better off under the decentralized system, even though it does not benefit from the transfers.

Where do we go from here? Fundamentally, there are two types of heterogeneity between regions that interact here: average income and group identity. In the absence of group loyalty, the interregional transfers that take place as a result of the uniformity under centralization are beneficial to everyone —the pooling effect. In other words, it is not the differences in regional average income *per se* that generate a cost under centralization, but rather the fact that together

[14]Notice, also, that it is so because we did not assume any other source of heterogeneity between regions. Indeed, had we assumed different regional weights for altruistic motives ($\alpha_A \neq \alpha_B$) or, alternatively, different median incomes ($y_A^R \neq y_B^R$), this would create such a political effect as in BR. That is, in the absence of transfers, the regions still have an incentive to secede as their preferences for redistribution differ.

with group loyalty, those differences might cause a decrease in the degree of solidarity in the society under direct democracy. Likewise, in the absence of interregional inequality, all individuals in both regions strictly prefer a centralized system of redistribution as long as group loyalty is not perfect —the scale effect. Again, it is not the non-internalization of spillovers that generate an inefficiency (i.e. an under-provision of redistribution) under decentralization when regions are identical, but rather the fact that the scale of redistribution is reduced from the point of view of the median voter.

As a result, then, the rationale for decentralization from an individual point of view —and thus from a welfare perspective under direct democracy— only arises when both sources of heterogeneity are present, that is, when there are both income disparities between regions and group loyalty.

Oates' (1972) decentralization theorem states that

1. *If there are no spillovers and regions are identical, then centralization and decentralization are equally efficient;*
2. *If there are no spillovers and regions are not identical, then decentralization is more efficient than centralization;*
3. *If there are spillovers and regions are identical, then centralization is more efficient than decentralization.*

In the traditional formulation of the above theorem, the term "spillover" relates to the fact that individuals value the presence of public goods in other jurisdictions to some extent. In that sense, the group loyalty parameter in our setup may be interpreted as the strength of regional spillovers regarding redistribution policy. That is, as long as group loyalty is not perfect (i.e. $\beta < 1$), individuals value —at least partly— both redistribution within and between regions.

Then, in the standard welfaristic approach from which the theorem follows, the cost of centralization comes from the uniform provision of public goods when regional preferences differ, whereas the cost of decentralization lies in the fact that local governments do not internalize the regional spillovers of the latter public goods. In the standard approach, then, the first-best solution depends on the level of spillovers, and is such that the provision of public goods differ across regions, provided that they have different preferences. Finally, the first-best solution can be reached whenever regions are identical and/or there are no spillovers.

In our political economy approach in the context of redistribution, in contrast, the cost of centralization from a welfare perspective lies in the fact that, in the presence of group loyalty, it may decrease the degree of solidarity in the society, which in turn is a consequence of the uniformity imposed by such a system[15]. Conversely, the welfare cost of decentralization relates to the fact

[15]Lockwood (2008) shows that Oates' theorem does not hold under majority voting (as opposed to the traditional welfaristic approach), even when the other assumptions are satisfied (i.e. uniformity and spillovers). In particular, he shows that, depending on the identity of the median voter (with respect to the average voter), the theorem might

that such a system does not allow to reduce interregional inequality. In our setup, then, the first-best solution does not depend on the level of spillovers, and is such that perfect equality among individuals is reached no matter whether the regions are identical or not. Finally, the first-best solution is never reached, as voters are partly self-interested.

Those differences between the two approaches in mind, we may still derive some interesting insights by comparing the decentralization theorem with the corresponding predictions of our political economy model of redistribution with interregional inequality and group loyalty:

1. *If there are no spillovers ($\beta = 1$) and regions are identical ($n_A^P = n_B^P$), then centralization and decentralization are equally efficient;*

2. *If there are no spillovers ($\beta = 1$) and regions are not identical ($n_A^P \neq n_B^P$), then centralization is more efficient than decentralization provided that the poor region is decisive under centralization. Otherwise, it is ambiguous which institutional arrangement yields higher total welfare;*

3. *If there are spillovers ($\beta < 1$) and regions are identical ($n_A^P = n_B^P$), then centralization is more efficient than decentralization;*

4. *If spillovers are total ($\beta = 1/2$), centralization is always more efficient than decentralization, no matter whether the regions are identical or not.*

Points (1) and (3) are identical to the ones in Oates' theorem. However, from point (2), it turns out that even when there are no regional spillovers, it is not the case that a decentralized system always yields higher total welfare. To the contrary, the pooling of resources makes total welfare strictly higher under centralization in most cases. From point (4), even though spillovers are total and regions are not identical, implying a trade-off in the traditional approach, centralization always welfare-dominates decentralization. As we already pointed out, the fact that $\beta = 1/2$ makes the regional heterogeneity regarding average income irrelevant from an individual point of view (the pooling effect). In fact, it turns out that more heterogeneity need not even increase the relative attractiveness of decentralization from a welfare perspective. Indeed, when $n_A^P = n_B^P$ and $\beta = 1/2$, decentralization is Pareto-dominated (the scale effect), but as $t_A = t_B$, rich and poor end up with the same final consumption in both regions. In contrast, when $n_A^P \neq n_B^P$ (and thus ($t_A \neq t_B$)), centralization becomes even more attractive for $\beta = 1/2$, as the fact that final consumptions differ between regions constitute an additional welfare loss, which is avoidable under centralization[16].

be reversed. Here, as we assume that all the rich are identical, our focus is not on the identification of the decisive (median) voter. The reason why the theorem fails in our setup relates to the benefits of centralization in terms of interregional redistribution, and this independently of whether it is an individual or a social planner who takes the decisions.

[16]Of course, whether the relative attractiveness of decentralization increases with interregional inequality depends on β. If $\beta = 1$ and the rich region is decisive, the relative efficiency of decentralization is actually increasing in inequality, as the latter is really harmful under the centralized solution.

3.7 Decentralization with Voluntary Transfers

As the cost of decentralization lies in the fact that no transfers occur between regions, a natural question that arises is whether —for the cases where $W(t) > W(t_A, t_B)$— centralization still welfare-dominates decentralization when allowing the rich region to transfer part of its tax revenue to the poor region under the decentralized system. Such a transfer allows for some beneficial pooling of resources, while it gives the rich region the opportunity to control the amount to be transferred to the poor region —and thus such a system may be easier to implement politically than a fully centralized system. However, the possibility of a transfer also creates an incentive for the rich in the poor region to free-ride on the generosity of the rich one. Indeed, knowing that the local poor will receive a transfer from outside, the rich in the poor region may consequently be less willing to contribute to local redistribution. Therefore, it is not obvious whether a decentralized system with transfers welfare-dominates a centralized one.

Notation 2. *From here on, $X(t_A, t_B, \theta)$ refers to X under decentralization with transfers. Furthermore, $t_j(\theta)$ refers to the equilibrium tax rate in region $j = A, B$ under decentralization with transfers.*

Formally, suppose that after the regional tax rates have been implemented, the rich region (region A) is allowed to transfer some proportion of its tax revenue to the poor region (region B). Clearly, the rich region is willing to do so provided that its decisive (rich) voter cares about the other region, that is, provided that $\beta_A < 1$[17]. The individuals' budget constraints are now given by

$$c_A^R(t_A, t_B, \theta) = (1 - t_A)y^R$$

$$c_B^R(t_A, t_B, \theta) = (1 - t_B)y^R$$

$$c_A^P(t_A, t_B, \theta) = y^P + \frac{t_A y^R n_A^R \theta}{n_A^P}$$

$$c_B^P(t_A, t_B, \theta) = y^P + \frac{t_B n_B^R y^R}{n_B^P} + \frac{t_A y^R n_A^R (1 - \theta)}{n_B^P}$$

where θ is the proportion of the tax revenue in A which stays in the region, and is to be determined endogenously. We consider a two-stage game: in the first stage, the regional tax rates are implemented, and in the second stage, the median voter in the rich region chooses θ. Solving the game backwards, we first determine the choice of θ for given (t_A, t_B) before solving for the equilibrium tax rates.

[17] As we focus on the motivation of region A to transfer part of its resources to the poor in region B, we now consider separately the effect of group loyalty in the two regions (i.e. β_j, $j = A, B$) on the equilibrium (t_A, t_B, θ).

3.7.1 The First-Best Solution

Suppose the social planner has to choose (t_A, t_B, θ) so as to maximize total welfare. Solving the game backwards, we get

$$(1-\theta(t_A^*, t_B^*)) = \frac{n_A^P n_B^P (t_B^* - t_A^*) + n(n_B^P t_A^* - n_A^P t_B^*)}{n_A^R (n_A^P + n_B^P) t_A^*}$$

The reaction functions are strictly decreasing, and

$$t_A^* = \frac{(y^R - y^P)}{y^R} \frac{(n_A^P + n_B^P)}{N} = t_B^*$$

Substituting, the optimal transfer is given by

$$(1-\theta^*) = \frac{n}{n_A^R} \frac{\left(n_B^P - n_A^P\right)}{\left(n_A^P + n_B^P\right)}$$

Therefore, (t_A^*, t_B^*, θ^*) are such that

$$c_A^R(t_A^*, t_B^*, \theta^*) = c_B^R(t_A^*, t_B^*, \theta^*) = c_A^P(t_A^*, t_B^*, \theta^*) = c_B^P(t_A^*, t_B^*, \theta^*) = \bar{c}$$

As expected, the optimal solutions of centralization and decentralization with transfers system coincide, that is, total welfare is maximized when equality among individuals is achieved, both within and between regions. Notice, furthermore, that this solution indeed constitutes a first-best in this economy, since the social planner does not wish to discriminate between neither the rich nor the poor individuals according to their region, even though he's allowed to.

3.7.2 Equilibrium and Welfare under Direct Democracy

In the second stage of the game, the median (rich) voter in region A chooses θ given (t_A, t_B) so as to maximize $U_A^R(t_A, t_B, \theta)$. The corresponding first-order condition is given by

$$\alpha\beta_A \left[\frac{n_A^P}{n_A} u'\left(c_A^P\right) \frac{t_A n_A^R y^R}{n_A^P}\right] - \alpha\left(1-\beta_A\right)\left[\frac{n_B^P}{n_B} u'\left(c_B^P\right) \frac{t_A n_A^R y^R}{n_B^P}\right] = 0$$

which simplifies to

$$\beta_A u'\left(c_A^P\right) = \left(1-\beta_A\right) u'\left(c_B^P\right)$$

Assuming that $u(c) = \ln c$, we get the transfer to region B as a function of the regional tax rates:

$$(1-\theta(t_A, t_B)) = \frac{(1-2\beta_A) n_A^P n_B^P y^P + y^R\left[(1-\beta_A) n_A^R n_B^P t_A - \beta_A n_A^P n_B^R t_B\right]}{\left[\beta_A n_A^P + (1-\beta_A) n_B^P\right] n_A^R t_A y^R}$$

and it has the following properties[18]:

- $\frac{\partial(1-\theta)}{\partial t_A} > 0$ and $\frac{\partial(1-\theta)}{\partial t_B} < 0$
- $\frac{\partial(1-\theta)}{\partial \beta_A} < 0$
- $\frac{\partial(1-\theta)}{\partial(y^R - y^P)} > 0$
- $\frac{\partial(1-\theta)}{\partial n_A^P} < 0$ and $\frac{\partial(1-\theta)}{\partial n_B^P} \lesseqgtr 0$

The properties of the transfer are intuitive. As expected, there is a threshold value of group loyalty $\beta_4 < 1$ above which the transfer is negative. Clearly, an increase in group loyalty in the second period has a negative effect on the transfer to the poor region. Observe that the effect of an increase in the proportion of poor in the poor region has an ambiguous effect on the transfer. Indeed, as more poor in B benefit from the transfer, the median voter in the rich region would like to transfer more. However, precisely because the poor in B are more numerous, the median voter also realizes that the transfer has to be shared among more individuals (i.e. it is more expensive to help those poor), which induces him to decrease the size of the transfer.

Suppose $\beta_A < \beta_4$, so that the transfer is positive —otherwise, we are back to the decentralized case without transfers. Substituting for $\theta\,(t_A, t_B)$ into $U_A^R(t_A, t_B, \theta)$ and $U_B^R(t_A, t_B, \theta)$, and solving for the first stage of the game, we get the reaction functions $t_A(t_B)$ and $t_B(t_A)$. Furthermore, for $i, j = A, B$, $i \neq j$,

$$\frac{\partial t_i(t_j)}{\partial t_j} < 0$$

that is, the regional tax rates are strategic substitutes. As we just saw, if a higher tax rate is implemented in region B in the first period, region A decreases the size of the transfer in the second period. Therefore, as region A keeps a bigger proportion of its own tax revenue for the local poor, it can decrease the regional tax rate accordingly. Likewise, if t_A is higher, the rich voter in the poor region correctly foresees that the rich region will be willing to increase the transfer. As a result, he implements a lower tax rate in equilibrium.

Substituting for $t_A\,(t_B)$ and $t_B\,(t_A)$ into one another, we can finally solve for (t_A, t_B, θ) as a function of the exogenous parameters. Doing so, we obtain that in equilibrium, $\frac{\partial t_A}{\partial \beta_A} < 0$, and $\frac{\partial t_B}{\partial \beta_A} > 0$. If β_A increases in the first period, the median voter in B foresees that the transfer in the second period will be lower. As a result, he implements a higher tax rate. In contrast, an increase in β_B has an ambiguous effect on t_B: on the one hand, it induces the decisive voter to redistribute more, as for the case without transfers. On the other hand, the decisive voter also realizes that by doing so, he will provoke a decrease in the transfer in the second period, which induces him to

[18] See section 2 of the Appendix for the derivatives computations.

decrease the tax rate. Therefore, the total effect depends on the relative magnitude of those two effects.

When β_A increases, the decisive voter in the rich region wants to transfer a smaller share of the local tax revenue to the poor region in the second period. As a result, he implements a lower tax rate in the first period. Observe that the effect of an increase in β_B on t_A is also ambiguous, as it obviously depends on how t_B will react to this increase in the first place, which is also ambiguous.

Suppose there is no group loyalty (i.e. $\beta = 1/2$ in both regions). In this case, $t_A(\theta) > t_B(\theta)$, and the transfer is strictly positive and such that $c_A^P(t_A, t_B, \theta) = c_B^P(t_A, t_B, \theta) < c^P(t)$. We then have the following result:

Proposition 6. *Assume region A is decisive under centralization and $n_A^P < n_B^P$. In the absence of group loyalty, it holds that*

$$W(t) > W(t_A, t_B, \theta) > W(t_A, t_B)$$

When $\beta = 1/2$, region A, being richer, implements a higher tax rate in the first period, while transferring part of its tax revenue to the poor region in the second period, so that the poor in both regions end up with the same final consumption. Observe, though, that the poor consume strictly less than when redistribution is centralized. Indeed, the rich voter in A, as he cannot force the rich in B to contribute as much as he would like to, chooses to redistribute strictly less in equilibrium.

Furthermore, it turns out that both $t_B(\theta) < t$ and $t_B(\theta) < t_B$. That is, the rich in B, knowing that there will be a transfer, chooses to contribute strictly less to redistribution (i.e. there is free-riding). This is because the rich in B, even though he cares equally about both regions, is still partly self-interested. As he knows that the local poor will benefit from the transfer, the rich in B implements a strictly lower tax rate than he would in the absence of the transfer, or than he would under centralization.

Therefore, in the absence of group loyalty, decentralization with transfers yields less redistribution than centralization, and, moreover, produces inequality between the rich across regions (i.e. $c_A^R(t_A, t_B, \theta) < c_B^R(t_A, t_B, \theta)$). As a result, centralizing redistribution yields the best outcome from a welfare perspective, although allowing for the transfer under decentralization is welfare-increasing in that case.

Finally, when $\beta = 1/2$, it turns out that except from the rich in the poor region, all individuals reach their highest final utility level under the centralized system. Again, the rich in A, even though decentralization with transfers gives him the control over the amount to be transferred to region B, is strictly better off under centralization, since it allows him to force the rich in B to contribute more to redistribution. In contrast, the rich in B reaches his highest utility level under decentralization with transfers, as it gives him the opportunity to free-ride on the generosity of region A.

Although the algebra involved does not allow us to prove it analytically, simulations of the model show that for all values of β for which the transfer is strictly positive (i.e. $\beta < \beta_4$), total welfare is strictly higher under centralization than under decentralization with transfers (see Figure 3.4). In other words, the welfare cost of free-riding always more than compensates the harmful effect of group loyalty on total welfare under centralization.

Furthermore, depending on the parameters, it turns out that we can have either $W(t_A, t_B, \theta) > W(t_A, t_B)$ or $W(t_A, t_B, \theta) < W(t_A, t_B)$ (see Figure 3.4). That is, it is not even generally true that total welfare under decentralization increases when allowing for the transfer. Again, although the transfer allows for some beneficial pooling of national resources, it also gives rise to harmful free-riding behavior in the poor region.

Yet, the stronger group loyalty, the more likely that the rich in A is better off under decentralization with transfers, as it allows him to better control the amount of local resources going to the poor region than does the uniformity imposed by centralization. Therefore, if the parameters of the economy are such that $W(t_A, t_B, \theta) > W(t_A, t_B)$, decentralization with transfers may be easier to sustain politically, and still welfare-dominates full decentralization.

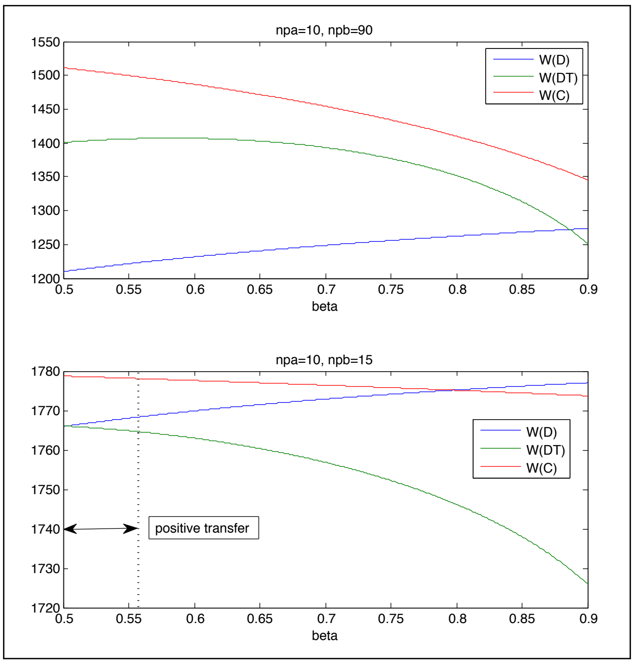

Figure 3.4: Total welfare under centralization and decentralization with and without transfers

3.8 Conclusion

We studied the choice between centralized and decentralized redistribution in a political economy context assuming both interregional income inequality and group loyalty. We showed that neither group loyalty alone, nor implicit interregional transfers alone, can rationalize the individual wish to decentralize. Whenever regions are equally rich, centralization always pareto-dominates decentralization as long as there are regional spillovers, since it makes the median voter more willing to redistribute as a result of such redistribution being implemented at a bigger scale. Likewise, whenever the two regions share a common identity, centralization always pareto-dominates decentralization, since it allows the median voter to pool national resources so as to homogenize redistribution in the country.

Implicit in our formulation is that decentralization has a cost in terms of interregional inequality. Always in the aggregate, but also from an individual perspective whenever group loyalty is not perfect. That is, in such case, individuals in the rich region want to transfer resources to the poor region to some extent, although the uniformity imposed by centralization makes it more and more likely that the rich region wishes to decentralize as long as group loyalty is getting stronger. In fact, if the rich region can unilaterally decide to decentralize redistribution, and if group loyalty is strong enough, the political equilibrium is very likely to be inefficient. Allowing the rich region to voluntarily transfer resources to the poor region under decentralization does not solve the problem, though: a centralized system is always better, as it can address the free-rider problem.

Observe that the rationale to decentralize here is a negative one: given that individuals in the rich region dislike interregional transfers, they are more willing to redistribute under decentralization, since regional taxes are then spent locally. If group loyalty is not too strong, however, there is some justification for a centrally implemented system of redistribution. Furthermore, decentralizing redistribution is likely to further increase interregional inequality in the long run, which may in turn increase group loyalty due to segregation and polarization.

In order to focus on donor motivation, and as the poor do not pay taxes in our set up, we have assumed that the median voter is always a rich individual. Furthermore, for tractability, we made the extreme simplifying assumption that all the rich and poor individuals have the same pre-tax income. In the spirit of Bolton and Roland (1997), a possible extension would be to assume instead a whole distribution of income within a region, while every voter pays taxes and receives the lump-sum transfer. Then, assuming that income distribution differs between regions (in the sense of first- or second-order stochastic dominance, for instance), it would be of interest to investigate how the identity of the median voter changes under a centralized system, and how this affects the choice between centralized and decentralized redistribution given the existence of group loyalty.

More generally, one could assume additional sources of heterogeneity between regions, such as the strength of altruism ($\alpha_A \neq \alpha_B$) or group loyalty ($\beta_A \neq \beta_B$). Intuitively, this would create more rationale for decentralization, as individual preferences for redistribution would now have regional specificities (preference-matching). For example, if region A is poorer (i.e. $n_A^P > n_B^P$) but values redistribution more than region B (i.e. $\alpha_A > \alpha_B$), it could well be the case that individuals in the poor region are better off under a decentralized system, if this allows them to redistribute more, even though they don't benefit from interregional transfers. A possibility is that the decentralized system with transfers would then dominate the centralized one. Indeed, the decentralized system would have the additional advantage of accommodating regional tastes regarding redistribution, while the transfer would allow to reduce interregional inequality. Of course, the free-riding problem would still exist, but it may not be true any longer that centralization always welfare-dominates decentralization with transfers.

A traditional argument against decentralized redistribution is the fact that with mobility, such a system creates tax competition between jurisdictions, and/or stratification of individuals by income. We have assumed away mobility of both tax payers and welfare recipients, as we believe that individual mobility motivated by tax/transfer differences between jurisdictions may be very limited in practice, especially in a context of cultural diversity, and given that groups are geographically segmented. However, a possible extension would be to assume imperfect mobility, and that the cost of moving for a given individual depends positively on the strength of group loyalty. For instance, if $\beta_i = 1/2$, there is no cost of moving, and if $\beta_i = 1$, an individual i never wants to move (and the cost is strictly positive for $1/2 < \beta_i < 1$). Intuitively, this could reinforce the fact that less group loyalty increases the relative efficiency of centralization under direct democracy. Indeed, the weaker group loyalty, the more the median voter is willing to redistribute under centralization, the more mobility under decentralization, and hence the (even) more attractive a centralized system of redistribution. Furthermore, if individuals are mobile, group loyalty would have a negative effect on redistribution under decentralization as well, which would clearly decrease the rationale for such a system.

3.9 References

1. Austen-Smith, D., and Wallerstein, M. (2006), Redistribution and Affirmative Action, *Journal of Public Economics*, Vol. 90, Issue 10-11, pp. 1789-1823.

2. Alesina, A., Baqir, R. and Easterly, W. (1999), Public Goods and Ethnic Divisions, *Quarterly Journal of Economics*, Vol. 114, Issue 4, pp. 1243-1284.

3. Besley, T. and Coate, S. (2003), Centralized versus Decentralized Provision of Local Public Goods: a Political Economy Approach, *Journal of Public Economics*, Vol. 87, Issue 12, pp. 2611-2637.

4. Bolton, P. and Roland, G. (1997), The Breakup of Nations: A Political Economy Analysis, *Quarterly Journal of Economics*, Vol. 112, No. 4, pp. 1057-1090.

5. Bjorvatn, K. and Cappelen, A. W. (2006), Redistribution and the Nature of Altruism: Should Welfare Programs Be Centralized or Decentralized?, *Economics of Governance*, Vol. 7, Issue 2, pp. 133-142.

6. Charness, G and Rabin, M. (2002), Understanding Social Preferences with Simple Tests, *Quarterly Journal of Economics*, Vol. 117, No. 3, pp. 817-869.

7. Collier, P. and Hoeffler, A (2006), The Political Economy of Secession, in Hurst Hannum and Eileen F. Babbitt (eds), *Negotiating Self-Determination*, Lanham: Lexington Books.

8. Fehr, E. and Schmidt, K. (1999), A Theory of Fairness, Competition, and Cooperation, *Quarterly Journal of Economics*, Vol. 114, No. 3, pp. 817-868.

9. Lind, J. T. (2004), Fractionalization and the Size of Government, *Journal of Public Economics*, Vol. 91, Issue 1-2, pp. 51-76.

10. Luttmer, E. F. P. (2001), Group Loyalty and the Taste for Redistribution, *Journal of Political Economy*, Vol. 109, Issue 3, pp. 500-528.

11. Lockwood, B. (2008), Voting, Lobbying, and the Decentralization Theorem, *Economics and Politics*, Vol. 20, No. 3, pp. 416-431.

12. Lockwood, B. (2002), Distributive Politics and the Costs of Centralization, *Review of Economic Studies*, Vol. 69, No. 2, pp. 313-337.

13. Meltzer, A. H. and Richard, S.F. (1981), A Rational Theory of the Size of Government, *Journal of Political Economy*, Vol. 89, No. 5, pp. 914-927.

14. Oates, W.E. (1972), Fiscal Federalism, New York: Harcourt Brace Jovanovich.

15. Pauly, M. V. (1973), Redistribution as a Local Public Good, *Journal of Public Economics*, Vol. 2, Issue 1, pp. 35-58.

16. Persson, T. and Tabellini, G. (1994), Does Centralization Increase the Size of Government?, *European Economic Review*, Vol. 38, Issue 3-4, pp. 765-773.8.

3.10 Appendix

3.10.1 Equilibrium Redistribution: Comparative Statics

The equilibrium tax rate in region $j = A, B$ under the *decentralized* solution is given by

$$t_j = \frac{\alpha\beta}{(1+\alpha\beta)} \frac{n_j^P}{n} \frac{\left(y^R - y^P\right)}{y^R} - \frac{1}{(1+\alpha\beta)} \frac{n_j^P}{n_j^R} \frac{y^P}{y^R}$$

Taking derivatives, we get

$$\frac{\partial t_j}{\partial n_j^P} = \frac{-\left[(1+\alpha\beta)n^2 - 2\alpha\beta n n_j^P + \alpha\beta(n_j^P)^2\right] y^P + \alpha\beta(n_j^R)^2 y^R}{(1+\alpha\beta)n(n_j^R)^2 y^R} > 0 \text{ if and only if } \alpha\beta > \frac{y^P}{(y^R - y^P)}\left(\frac{n}{n_j^R}\right)^2$$

$$\frac{\partial t_j}{\partial y^R} = \frac{n_j^P(n+\alpha\beta n_j^R)y^P}{(1+\alpha\beta)n n_j^R (y^R)^2} > 0 \text{ and } \frac{\partial t_j}{\partial y^P} = -\frac{n_j^P(n+\alpha\beta n_j^R)}{(1+\alpha\beta)n n_j^R y^R} < 0$$

$$\frac{\partial t_j}{\partial \alpha} = \frac{\beta n_j^P(n_j^P y^P + n_j^R y^R)}{(1+\alpha\beta)^2 n n_j^R y^R} > 0 \text{ and } \frac{\partial t_j}{\partial \beta} = \frac{\alpha n_j^P(n_j^P y^P + n_j^R y^R)}{(1+\alpha\beta)^2 n n_j^R y^R} > 0$$

The equilibrium tax rate under the *centralized* solution is given by

$$t = \frac{1}{(1+\alpha)} \left\{ \alpha \left[\beta \frac{n_A^P}{n} + (1-\beta) \frac{n_B^P}{n} \right] - \frac{y^P}{y^R} \frac{(n_A^P + n_B^P)}{(n_A^R + n_B^R)} \left[1 + \alpha\beta \frac{n_A^R}{n} + \alpha(1-\beta) \frac{n_B^R}{n} \right] \right\}$$

Taking derivatives, we get

$$\frac{\partial t}{\partial n_A^P} \gtreqless 0 \text{ if and only if } \beta \gtreqless \beta_1 = \frac{2n\left[n_A^P + n_A^R(1+\alpha)\right] y^P}{\alpha y^P \left[4nn_A^P - (n_A^P + n_B^P)^2\right] + \alpha y^R (n_A^R + n_B^R)^2}$$

$$\frac{\partial t}{\partial n_B^P} \gtreqless 0 \text{ if and only if } \beta \lesseqgtr \beta_2 = \frac{\left[2(1+\alpha)n^2 + \alpha(n_A^P + n_B^P)^2 - 2\alpha n(n_A^P + 2n_B^P)\right] y^P - \alpha y^R (n_A^R + n_B^R)^2}{\alpha y^P \left[-4nn_B^P + (n_A^P + n_B^P)^2\right] - \alpha y^R (n_A^R + n_B^R)^2}$$

$$\frac{\partial t}{\partial n_B^P} = -\frac{2\left(n + \alpha n_A^R\right) y^P}{(1+\alpha)\left(n_A^R + n_B^R\right)^2 y^R} < 0 \text{ if } \beta = 1$$

$$\frac{\partial t}{\partial y^R} = \frac{(n_A^P + n_B^P)\left[n + \alpha n_B^R + \alpha\beta(n_B^P - n_A^P)\right] y^P}{(1+\alpha) n (n_A^R + n_B^R)(y^R)^2} > 0$$

$$\frac{\partial t}{\partial y^P} = \frac{\left(n_A^P + n_B^P\right)\left\{\alpha\left[\beta\left(n_A^P - n_B^P\right) + n_B^P\right] - (1+\alpha)n\right\}}{(1+\alpha) n \left(n_A^R + n_B^R\right) y^R} < 0$$

$$\frac{\partial t}{\partial \alpha} = \frac{\left[\beta(n_A^P - n_B^P) + n_B^P\right]\left[(n_A^P + n_B^P) y^P + (n_A^R + n_B^R) y^R\right]}{(1+\alpha)^2 n (n_A^R + n_B^R) y^R} > 0$$

$$\frac{\partial t}{\partial \beta} = \frac{\alpha}{(1+\alpha)} \frac{\left(n_A^P - n_B^P\right)}{n} \left[1 + \frac{\left(n_A^P + n_B^P\right) y^P}{\left(n_A^R + n_B^R\right) y^R}\right] > 0 \text{ if and only if } n_A^P > n_B^P$$

3.10.2 Decentralization with Transfers: Comparative Statics

Assuming that $u(c) = \ln c$, the transfer to region B is given by

$$(1-\theta) = \frac{(1 - 2\beta_A) n_A^P n_B^P y^P + y^R \left[(1-\beta_A) n_A^R n_B^P t_A - \beta_A n_A^P n_B^R t_B\right]}{n_A^R \left[\beta_A \left(n_A^P - n_B^P\right) + n_B^P\right] t_A y^R}$$

Taking derivatives,

$$\frac{\partial (1-\theta)}{\partial t_A} = \frac{(2\beta_A - 1) n_A^P n_B^P y^P + y^R t_B \beta_A n_A^P n_B^R}{n_A^R \left[\beta_A \left(n_A^P - n_B^P\right) + n_B^P\right] (t_A)^2 y^R} > 0$$

$$\frac{\partial (1-\theta)}{\partial t_B} = -\frac{\beta_A n_A^P n_B^R}{n_A^R \left[\beta_A \left(n_A^P - n_B^P\right) + n_B^P\right] t_A} < 0$$

$$\frac{\partial (1-\theta)}{\partial \beta_A} = -\frac{n_A^P n_B^P \left[n_B (t_A + t_B) y^R + n_A^P \left(y^P - t_A y^R\right) + n_B^P \left(y^P - t_B y^R\right)\right]}{n_A^R \left[\beta_A \left(n_A^P - n_B^P\right) + n_B^P\right]^2 t_A y^R} < 0$$

$$\frac{\partial (1-\theta)}{\partial y^R} = \frac{(2\beta_A - 1) n_A^P n_B^P y^P}{n_A^R \left[\beta_A \left(n_A^P - n_B^P\right) + n_B^P\right] t_A \left(y^R\right)^2} > 0$$

$$\frac{\partial (1-\theta)}{\partial y^P} = \frac{(1 - 2\beta_A) n_A^P n_B^P}{n_A^R \left[\beta_A \left(n_A^P - n_B^P\right) + n_B^P\right] t_A y^R} < 0$$

$$\frac{\partial(1-\theta)}{\partial n_A^P} = \frac{\begin{array}{c}(2\beta_A - 1)\, n_B^P \left[-\beta_A \left(n_A^P\right)^2 + (\beta_A - 1)\, n_B n_B^P\right] y^P \\ +\beta_A \left\{(\beta_A - 1)\left(n_A^R\right)^2 n_B^P t_A + n_B^R \left[-\beta_A \left(n_A^P\right)^2 + (\beta_A - 1)\, n_B n_B^P\right] t_B\right\} y^R\end{array}}{\left(n_A^R\right)^2 \left[\beta_A\left(n_A^P - n_B^P\right) + n_B^P\right]^2 t_A y^R} < 0$$

$$\frac{\partial(1-\theta)}{\partial n_B^P} = \frac{\beta_A n_A^P \left\{-(\beta_A - 1)\, n\, (t_A + t_B)\, y^R + n_A^P \left[y^P - 2\beta_A y^P - t_A y^R + \beta_A (t_A + t_B)\, y^R\right]\right\}}{n_A^R \left[\beta_A\left(n_A^P - n_B^P\right) + n_B^P\right]^2 t_A y^R} \lesseqgtr 0$$

Taking derivatives of the equilibrium values of (t_A, t_B), we get

$$\frac{\partial t_A}{\partial \beta_A} = -\frac{\begin{array}{c}\left[n + \alpha\beta_B n - \alpha\,(\beta_B - 1)\, n_A^P\right]\left[(\beta_B - 1)\, n_A^P - \beta_B n_B^P\right] \\ \left[-\alpha n_A^P n_B^P + n\left(-n_A^P + n_B^P + \alpha n_B^P\right)\right]\left[\left(n_A^P + n_B^P\right) y^P + \left(n_A^R + n_B^R\right) y^R\right]\end{array}}{n_A^R \left\{\alpha(\beta_A + \beta_B - 1) n_A^P n_B^P - n\left[(1 + \beta_A + \alpha\beta_A - \beta_B) n_A^P + (1 - \beta_A + \beta_B + \alpha\beta_B) n_B^P\right]\right\}^2 y^R} < 0$$

$$\frac{\partial t_B}{\partial \beta_A} = \frac{\begin{array}{c}\left(n + \alpha\beta_B n_B^R\right)\left[n_A^P - \beta_B n_A^P + \beta_B n_B^P\right] \\ \left[-\alpha n_A^P n_B^P + n\left(-n_A^P + n_B^P + \alpha n_B^P\right)\right]\left[\left(n_A^P + n_B^P\right) y^P + \left(n_A^R + n_B^R\right) y^R\right]\end{array}}{n_B^R \left\{\alpha(\beta_A + \beta_B - 1) n_A^P n_B^P - n\left[(1 + \beta_A + \alpha\beta_A - \beta_B) n_A^P + (1 - \beta_A + \beta_B + \alpha\beta_B) n_B^P\right]\right\}^2 y^R} > 0$$

3.10.3 Proofs

Proof of Proposition 1. Assume that $n_A^P \neq n_B^P$. It follows that $W(t_A^*, t_B^*) > W(t^*)$ if and only if

$$n_A^R \ln \frac{c_A^R(t_A^*, t_B^*)}{c^R(t^*)} + n_A^P \ln \frac{c_A^P(t_A^*, t_B^*)}{c^P(t^*)} + n_B^R \ln \frac{c_B^R(t_A^*, t_B^*)}{c^R(t^*)} + n_B^P \ln \frac{c_B^P(t_A^*, t_B^*)}{c^P(t^*)} > 0$$

We know that t_j^* is such that $c_j^P(t_A^*, t_B^*) = c_j^R(t_A^*, t_B^*) = \bar{c}_j$, $j = A, B$. Furthermore, t^* is such that $c^P(t^*) = c^R(t^*) = \bar{c}$ in both regions. Therefore, $W\left(t_A^*, t_B^*\right) > W\left(t^*\right)$ if and only if $\bar{c}_B \bar{c}_A > \bar{c}^2$, or, equivalently, $(1 - t_A^*)(1 - t_B^*) > (1 - t^*)^2$. Substituting, this condition becomes

$$\left[\frac{(n_A^P - n_B^P)(y^R - y^P)}{2N}\right]^2 < 0$$

which never holds. Therefore, total welfare is strictly higher under the centralized solution. □

Proof of Proposition 3. Assume $n_A^P = n_B^P = n^P$. In that case, we know that $U_A^i(t_A, t_B) = U_B^i(t_A, t_B) = U^i(t_A, t_B)$ and $U_A^i(t) = U_B^i(t) = U^i(t)$, $i = R, P$ (i.e. utility levels are the same across regions under both centralization and decentralization). A voter i is better off under decentralization if and only if $U^i\left(t_A, t_B\right) > U^i\left(t\right)$, $i = R, P$. When $n_A^P = n_B^P$ and $\beta = 1$, we know that $t_A = t_B = t$, and since there are no interregional transfers taking place under centralization, it follows that $U^i\left(t_A, t_B\right) = U^i\left(t\right)$, that is, all voters in the economy are indifferent between centralized and decentralized redistribution. Then, as $\frac{\partial U^i(t)}{\partial \beta} = 0$, $i = R, P$,

$$\frac{\partial U^P(t_A, t_B)}{\partial \beta} = \frac{n^P(1 + \alpha) + n^R(1 + \alpha\beta)}{\beta(1 + \alpha\beta)\left[n^P + (1 + \alpha\beta)n^R\right]} > 0$$

and

$$\frac{\partial U^R(t_A, t_B)}{\partial \beta} = \frac{\alpha(1-\beta)n^P}{\beta(1+\alpha\beta)\left[n^P + (1+\alpha\beta)n^R\right]} > 0$$

it follows directly that all voters are strictly better off under centralized redistribution whenever $\beta < 1$. □

Proof of Proposition 4. In order to prove Proposition 4, we will use the solution of an intermediary institutional arrangement between centralization and decentralization, namely decentralization with voluntary transfers, which we analyze in section 7 in the main text.

Suppose that after the regional tax rates have been implemented, the rich region (region A) is allowed to transfer some proportion of its tax revenue to the poor region (region B). The individuals' budget constraints are now given by

$$c_A^R(t_A, t_B, \theta) = (1 - t_A)y^R$$

$$c_B^R(t_A, t_B, \theta) = (1 - t_B)y^R$$

$$c_A^P(t_A, t_B, \theta) = y^P + \frac{t_A y^R n_A^R \theta}{n_A^P}$$

$$c_B^P(t_A, t_B, \theta) = y^P + \frac{t_B n_B^R y^R}{n_B^P} + \frac{t_A y^R n_A^R (1-\theta)}{n_B^P}$$

where θ is the proportion of the tax revenue in A which stays in the region, and is to be determined endogenously. We consider a two-stage game: in the first stage, the regional tax rates are implemented, and in the second stage, the median voter in the rich region chooses θ. Solving the game backwards, we first determine the choice of θ for given (t_A, t_B), after which we solve for the equilibrium regional tax rates.

In the second stage of the game, the median (rich) voter in region A chooses θ given (t_A, t_B) so as to maximize $U_A^R(t_A.t_B, \theta)$. The corresponding first-order condition is given by

$$\beta u'(c_A^P) = (1-\beta)u'(c_B^P)$$

Assuming that $u(c) = \ln c$, we get the transfer as a function of the regional tax rates:

$$(1 - \theta(t_A, t_B)) > 0 \text{ if and only if } \beta < \frac{n_B^P(n_A^P y^P + n_A^R t_A y^R)}{2n_A^P n_B^P y^P + y^R(n_A^R n_B^P t_A + n_B^R n_A^P t_B)} = \beta_4 < 1$$

Assume $\beta < \beta_4$, so that the transfer is positive —otherwise, we are back to the decentralized case without transfer. Substituting for $\theta(t_A, t_B)$ into $U_A^R(t_A, t_B, \theta)$ and $U_B^R(t_A, t_B, \theta)$, and solving for the first stage of the game, we get the reaction functions $t_A(t_B)$ and $t_B(t_A)$. Furthermore, for $i, j = A, B$, $i \neq j$,

$$\frac{\partial t_i(t_j)}{\partial t_j} < 0$$

that is, the regional tax rates are strategic substitutes. Then, substituting for t_A (t_B) and t_B (t_A) into one another, we can finally solve for (t_A, t_B, θ) as a function of the exogenous parameters.

Suppose now that a social planner has to choose (t_A, t_B, θ) so as to maximize $W(t_A, t_B, \theta)$. Solving the game backwards, we have that $t_A^* = t_B^*$, and θ^* is such that

$$c_A^R(t_A^*, t_B^*, \theta^*) = c_B^R(t_A^*, t_B^*, \theta^*) = c_A^P(t_A^*, t_B^*, \theta^*) = c_B^P(t_A^*, t_B^*, \theta^*) = \overline{c}$$

That is, the welfare-maximizing solution of centralization and decentralization with transfers coincide, and are such no inequality between any two individuals remains. Therefore, since under both regimes, welfare is increasing in general redistribution until achieving equality of final consumption between individuals, it follows that we can use total variance as a criterion to compare both systems from a welfare perspective, total variance being defined by

$$V = \frac{1}{N} \sum_{j=A,B} \sum_{i=R,P} n_j^i (c_j^i - \overline{c})^2$$

Now suppose that $\beta = 1/2$. Then, $V(t_A, t_B, \theta) - V(t)$ is given by

$$\frac{1}{N^2} \left\{ \left[-\frac{(n_A^P + n_B^P)}{(n_A^R + n_B^R)(1+\alpha)^2} + \frac{2\left[-2n_A^P n_B^P + n(n_A^P + n_B^P)\right]}{n_A^R n_B^R (2+\alpha)^2} \right] \left[(n_A^P + n_B^P) y^P + (n_A^R + n_B^R) y^R\right]^2 \right\} > 0$$

from which it follows that $W(t) > W(t_A, t_B, \theta)$.

Furthermore, for $\beta = 1/2$, $W(t_A, t_B, \theta) - W(t_A, t_B)$ is given by

$$(1+\alpha) n \ln \left[\frac{\left[(n_A^P + n_B^P) y^P + (n_A^R + n_B^R) y^R\right]^2}{4(n_A^P y^P + n_A^R y^R)(n_B^P y^P + n_B^R y^R)} \right] > 0$$

and so, by transitivity, $W(t) > W(t_A, t_B)$ in the absence of group loyalty.

It remains to be proven that $U_j^i(t) > U_j^i(t_A, t_B)$ for $i = R, P$ and $j = A, B$. Given that $\beta = 1/2$, we have that

$$t_A(\theta) - t_B(\theta) = \frac{\left(n_B^P - n_A^P\right) \left[\left(n_A^P + n_B^P\right) y^P + \left(n_A^R + n_B^R\right) y^R\right]}{(2+\alpha)\, n_A^R n_B^R y^R} > 0$$

and the transfer is positive and such that

$$c_A^P(t_A, t_B, \theta) = \frac{\alpha}{(2+\alpha)} \left[\frac{(n_A^P + n_B^P) y^P + (n_A^R + n_B^R) y^R}{N} \right] = c_B^P(t_A, t_B, \theta)$$

Comparing the centralized solution with the decentralized solution with transfer when $\beta = 1/2$ we get

$$c^P(t) - c^P(t_A, t_B, \theta) = \frac{\alpha}{(2+\alpha)(1+\alpha)} \left[\frac{(n_A^P + n_B^P) y^P + (n_A^R + n_B^R) y^R}{N} \right] > 0$$

And comparing centralization and decentralization when $\beta = 1/2$,

$$c^P(t) - c^P_B(t_A, t_B) = \frac{\alpha}{2(1+\alpha)(2+\alpha)} \left[2y^R n^R_A + \alpha(n^P_B - n^P_A)(y^R - y^P)\right] > 0 \text{ whenever } n^P_A < n^P_B$$

Recall that when $\beta = 1/2$, the final utility of any voter $i = R, P$ in region $j = A, B$ is an increasing function of total welfare under the corresponding solution:

$$U^i_j = u(c^i_j) + \frac{\alpha}{(1+\alpha)} \frac{W}{N} \tag{3.5}$$

As a result, the preferred tax rate and final utilities of the rich and poor in A and B under centralization coincide. Hence, we can assume without loss of generality that $n^P_A < n^P_B$.

- Given that $W(t) > W(t_A, t_B)$ and $c^P_B(t) > c^P_B(t_A, t_B)$, it follows directly from (5) that $U^P_B(t) > U^P_B(t_A, t_B)$.
- Consider the rich in A. Given that $\theta < 1$ (i.e. the transfer is positive), it follows directly that $U^R_A(t_A, t_B, \theta) > U^R_A(t_A, t_B)$. Otherwise, the rich in A would just set $\theta = 1$ (i.e. no transfer), so that $U^R_A(t_A, t_B, \theta) = U^R_A(t_A, t_B)$. In other words, the rich in A is better off when resources can be transferred to region B. But then, given that $c^P(t) > c^P(t_A, t_B, \theta)$, it is also direct that $U^R_A(t) > U^R_A(t_A, t_B, \theta)$. Indeed, the rich (decisive) voter in A could well implement $\widetilde{t} < t$ under centralization such that $c^P(\widetilde{t}) = c^P(t_A, t_B, \theta)$. In this case, it would hold that $U^R_A(\widetilde{t}) > U^R_A(t_A, t_B, \theta)$, since the rich decisive voter in A would force the all the rich in B to contribute to redistribution as much as he wants, while the poor end up with the same consumption. Finally, since t maximizes $U^R_A(t)$, we have that $U^R_A(t) > U^R_A(\widetilde{t})$, and hence, by transitivity, $U^R_A(t) > U^R_A(t_A, t_B)$.
- If $U^R_A(t) > U^R_A(t_A, t_B)$, then, it follows directly that $U^R_B(t) > U^R_B(t_A, t_B)$. Indeed, given the direction of the interregional transfers under centralization (i.e. from A to B), there is no rationale for the rich in B to prefer decentralization provided that the rich in A is better off under centralization.
- Consider finally the poor in A. For $\beta = 1/2$, we have that

$$\frac{c^P(t)}{c^P_A(t_A, t_B)} - \frac{c^R(t)}{c^R_A(t_A, t_B)} = \frac{(2+\alpha)nn^R_B\left[(n^P_A + n^P_B)y^P + (n^R_A + n^R_B)y^R\right]}{(1+\alpha)\left[(2+\alpha)n - \alpha n^P_A\right](n^R_A + n^R_B)(n^P_A y^P + n^R_B y^R)} > 0$$

  This, in turn, implies that $U^P_A(t) - U^P_A(t_A, t_B) > U^R_A(t) - U^R_A(t_A, t_B) > 0$, and so, by transitivity, $U^P_A(t) - U^P_A(t_A, t_B) > 0$.

Therefore, centralization pareto-dominates decentralization in the absence of group loyalty.

□

Proof of Proposition 5. As the decisive voter is partly self-interested, he always implements a tax rate such that the rich consume strictly more than the poor under both centralization and decentralization. Indeed, we have that

$$c^R(t) - c^P(t) = \frac{\left\{ [1+\alpha(1-2\beta)]\, n_A^P + [1-\alpha(1-2\beta)]\, n_B^P \right\} \left[(n_A^P + n_B^P) y^P + (n_A^R + n_B^R) y^R \right]}{(1+\alpha)(n_A^P + n_B^P)(n_A^R + n_B^R)} > 0$$

$$c_j^R(t_A, t_B) - c_j^P(t_A, t_B) = \frac{n_j^P y^P + n_j^R y^R}{(1+\alpha\beta) n_j^R} > 0,\ j = A, B$$

As total welfare under decentralization is maximized for t_j^* such that $c_j^R = c_j^P$, $j = A, B$, and as

$$\frac{\partial W(t_i, t_j)}{\partial t_i} = (1+\alpha) y^R \left[\frac{n_i^R n_i^P}{n_i^P y^P + n_i^R t_i y^R} - \frac{n_i^R}{(1-t_i) y^R} \right] > 0 \text{ for } t_i < t_i^*,\ i, j = A, B$$

the result follows directly from the fact that $\frac{\partial t_j}{\partial \beta} > 0$, $j = A, B$. Similarly, as total welfare under centralization is maximized for t^* such that $c^R = c^P$, and as

$$\frac{\partial W(t)}{\partial t} = (1+\alpha) y^R \left[\frac{n_A^R n_A^P}{n_A^P y^P + n_A^R t y^R} + \frac{n_B^R n_B^P}{n_B^P y^P + n_B^R t y^R} - \frac{n_A^R + n_B^R}{(1-t) y^R} \right] > 0 \text{ for } t < t^*$$

the result follows directly from the fact that $\frac{\partial t}{\partial \beta} > 0$ if and only if $n_A^P > n_B^P$. □

Proof of Proposition 6. See the proof of Proposition 4. □

Chapter 4

Decentralization as a Way to Avoid Secessionist Conflict: the Role of Interregional Inequality

4.1 Introduction

Nation boundaries have been in movement for a long time. While the emergence of some countries has in some cases been the result of peaceful separations (like the separation of Slovenia from Yugoslavia), in many other cases separatist conflicts have shown a high degree of violence (like Bosnia, Croatia, or Pakistan and Bangladesh). Furthermore, secessionist movements are currently present in several countries (like Belgium, Spain or Canada). According to Gurr (2000), such secessionist movements were present in at least 52 countries between 1985 and 1999.

In this context, decentralization is often seen as a way of accommodating regional demands for more autonomy in heterogeneous countries. To the extent that decentralizing allows regional governments to implement policies that are closer to local preferences, it potentially constitutes an alternative to costly and sometimes violent secession attempts. Yet, decentralization does better in curbing secessionism in some countries (such as Canada, Spain, or Switzerland) than in others (such as Yugoslavia, Czechoslovakia or Indonesia) (Lake and Rotchild (2005)).

Decentralization, although it brings local governments "closer to the people", also has the tendency to exacerbate interregional inequality (Linz and Stepan (2000)), and it might do so through several mechanisms (Bakke and Wibbels (2006)). First, in a highly decentralized setting, the fiscal tools at the disposal of the central government to engage in redistribution between regions might be quite diminished. Second, decentralized intergovernmental competition for capital can exacerbate inequalities, as poor regions have little potential to attract capital and rich regions actually draw capital out of poor regions. Third, the tendency for subnational governments to act as veto players at the national level in fiscally decentralized settings might prevent the implementation of extensive interregional redistributive policies by the central government. According to Bakke and Wibbels (2006, p. 17), "*the net result of these factors may be that wealthy regions are able to fund substantial provision of public goods, crowd in private sector investment, and grow relatively quickly while poor regions lag ever farther behind*".

In this paper, we focus on the latter particular channel through which decentralization might actually contribute to conflict, namely interregional inequality. If one region is poorer than the other, decentralization, by reducing the extent of interregional redistribution, and thereby exacerbating inequality, might actually fuel conflict between regions rather than appeasing it. That is, if decentralization theoretically allows the regions to implement their preferred policies, it is of little help if at the same time, it makes some of them loose the resources necessary to do it. As a result, the relevance of decentralization as a conflict-mitigating strategy could well depend on the underlying structure of inequality in the country. In particular, we argue that if interregional inequality is important, decentralization, by fostering resentment in poorer regions, is very likely to contribute to conflict.

As we aim at analyzing whether separatist conflict can be avoided using partial decentralization, this means that we consider such decentralization as an *alternative* to conflict. In other words, we do not model the link between conflict and decentralization in a direct manner, that is, we do not assume that decentralization can directly affect the intensity of a given conflict. Rather, we ask whether it can serve as a peaceful device in order to avoid conflict, and in particular, we ask whether the answer to this question depends on the extent of inequality between regions.

Finally, observe that our reasoning implies a double-sided relationship between inequality and decentralization. First, as we said, decentralization has the tendency to increase interregional inequality, which might contribute to conflict. But then, since decentralization is used precisely to avoid such conflict, the latter implies that it should be limited in the presence of high inequality between regions. That is, decentralization is endogenous to the territorial structure of inequality. Therefore, we have that decentralization increases inequality, but also that inequality, in turn, conditions the level of decentralization in the shadow of conflict. This is consistent with the fact that *"the empirical evidence is not conclusive with regards to the exact nature and direction of causality of the relationship between fiscal decentralization and inequality, nor on the sign of the relationship itself"* (Sacchi and Salotti (2011, p. 6)).

We set up a simple political economy model of nation formation in the spirit of Alesina and Spolaore (1997), in which voters have to choose where to locate a public good and suffer disutility in terms of the distance between the public good and their ideal point. Furthermore, we assume that direct redistribution between regions takes place under unification, which is also the status quo. We then assume that there exists a conflict technology which allows the regions to fight in order to force the border configuration they prefer. In this context, we show that interregional inequality fuels conflict in both the poor and the rich region. However, and interestingly, more interregional inequality may result in either a lower or a higher probability of a secession occurring through conflict in equilibrium, depending on whether unifying is efficient or not.

The relative efficiency of unifying versus seceding, in turn, depends on the resolution of the traditional trade off between preferences heterogeneity and the existence of economies of scale in the production of the public good. If the former effect dominates, seceding is the socially efficient outcome (and thus we have a conflict in an inefficient union), whereas if the latter effect dominates, unifying is the socially efficient outcome (and thus we have conflict in an efficient union).

The key element determining the intensity of conflict in the two regions is their relative stakes in such conflict. In the setup we just described, there is only one possible conflict configuration, namely, that the rich region seeks secession, while the poor region seeks unification. More specifically, the higher inequality for a given level of redistribution, the more potential disagreement there is between the two regions regarding the choice of border configuration. Indeed, both the preferences heterogeneity and economies of scale effects make the two regions' interest on that matter go in the same direction, while it is the opposite as far as inequality and redistribution are concerned. Indeed, if there are higher benefits from economies of scale, both regions find unifying relatively more attractive, whereas if inequality increases, the rich region has more incentives to secede, while the poor region conversely has more incentives to preserve the union. Therefore, if inequality and/or redistribution is high enough as compared to the benefits from unifying, the rich region starts a conflict in order to secede from the union.

Given that the rich region fights in order to secede, while the poor region fights in order to resist it, it follows that the regional stakes in conflict are asymmetric. Which of the two regions devotes more resources to conflict, in turn, depends on whether unifying is more efficient than seceding with respect to the traditional trade off between economies of scale and heterogeneity of preferences. Suppose that the economies of scale effect dominates the heterogeneity effect, or, in other words, the union between the two regions is efficient. Given that the poor region fights in order to preserve the union, this means that it has more to win from fighting than the rich region, that is, the poor region has relatively bigger stakes in the conflict. As a result, the poor region devotes strictly more resources to such conflict than does the rich region.

The relative size of regional conflict efforts, in turn, depends on the extent of such asymmetry in regional stakes. In an efficient union, the benefits from economies of scale more than compensate the costs from heterogeneity. The bigger those economies of scale, then, the bigger the relative efficiency of the union as compared to a secession, meaning that the stakes are correspondingly more asymmetric. Indeed, the poor region has even more to win from resisting a secession attempt, whereas the rich region has relatively less to win from such a potential secession. As a result, the relative size of conflict effort in the poor region increases, and its winning probability increases correspondingly. Furthermore, it turns out that total conflict intensity is also directly related to the regional stakes' asymmetry. Specifically, the bigger this asymmetry (i.e. the bigger the economies of scale in an efficient union, or the bigger the heterogeneity costs in an inefficient union), the less intense is the conflict. In short, the bigger the asymmetry regarding the regional stakes in conflict, the less intense the conflict, and the more certain its outcome[1].

In this context, it turns out that an increase in interregional inequality always has the effect of bringing the regional stakes in conflict closer to each other. Indeed, no matter whether unifying is efficient or not, more inequality always translates into bigger stakes for *both* regions in the conflict, through interregional redistribution. It follows that the bigger inequality, the bigger its relative importance with respect to the public good from an utility perspective in both regions. Therefore, since more inequality implies more symmetry in regional stakes, by the same reasoning we just described, it is always associated with higher conflict intensity, together with a closer relative size of regional conflict efforts. Now, whether as a result of such an increase in inequality, a secession becomes more or less likely, depends on which region is willing to invest more in conflict in the first place, and thus on whether unifying is efficient or not. In any case, however, more inequality always brings the winning probabilities closer to each other (i.e. the conflict outcome is more uncertain).

Interregional inequality is thus very harmful for welfare, not because of inequality aversion *per se*, but because it fuels wasteful conflict in both regions, while decreasing the probability of the most efficient outcome. Therefore, the question we aim at answering is whether a peaceful compromise can be reached by the use of partial decentralization. That is, we ask whether there exists an intermediate solution between unification and secession, which we call partial decentralization, such that in both regions, individuals are better off under this alternative than under the conflictual outcome. If so, we will say that the peaceful —decentralized— outcome is politically sustainable.

Assuming that partial decentralization applies to both interregional redistribution and the supply of public goods, we show that, as a result of conflict being totally wasteful, there always exists a range of decentralization levels such that peace is politically sustainable. Then, we show that a rise in interregional inequality has two effects on this range: first, both the minimum and the maximum levels of decentralization compatible with peace may be either smaller or bigger depending on whether unifying is efficient, and second, the range of peace-sustainable decentralization levels is strictly bigger. That is, the higher inequality, the more room there is for a peaceful compromise.

Given that a range of decentralization levels compatible with peace exists for all possible levels of inequality, we then show that the peaceful outcome is always implemented in a direct democracy. That is, conflict never occurs in equilibrium, since individuals in both regions prefer to implement a partial decentralization level compatible with peace rather than starting a conflict. In any case, whatever region is decisive on that matter, the level of partial decentralization that is implemented does depend on interregional inequality. In particular, in an efficient union, the level of decentralization that is implemented in order to avoid conflict is increasing in inequality, while the opposite holds in an inefficient union.

[1] The fact that conflict effort increases with the symmetry of stakes is similar to previous findings according to which symmetry in competitive advantage (capability) tends to enhance individual performance. See for instance Lazear and Rosen (1981), who show that a handicapping system induces efficient competition in a rank-order tournament between weak and strong players, or Myerson (2001), who shows that revenue maximizing auction between asymmetric bidders implies favoring weak bidders.

Provided that partial decentralization is implemented and serves as a peaceful compromise between the interests of the two regions, the question, in turn, is whether there are no incentives left to start a conflict under the decentralized outcome. That is, we ask whether peace is self-enforcing. It turns out that whether unification is still an option once decentralization has been implemented constitutes the crucial element in order to answer this question. We show that, if this is not the case, peace is not self-enforcing, and no partial decentralization is ever implemented. As a result, when it is an irreversible process, partial decentralization cannot serve as a mean to achieve peace, and separatist conflict always occurs in equilibrium.

Whether the conflictual outcome welfare-dominates the peaceful one that is reached through partial decentralization, in turn, depends on the level of inequality between regions. As more inequality translates into more conflict intensity in both regions, together with a lower probability of the efficient outcome, it follows that more inequality also makes the peaceful solution more attractive from a welfare perspective. In fact, when inequality between regions is high enough, conflict becomes so costly that peace welfare-dominates conflict for all partial decentralization levels.

Our analysis relates to the recent literature on the political economy of nation formation. Goyal and Staal (2004), using the same type of model, study the effect of size and strategic location of regions on the incentives to secede. However, they do not allow for conflict nor for the possibility of partially decentralizing public policy. The effect of income differences between regions on the incentives to secede has been analyzed by Bolton and Roland (1997). They focus on redistribution policy, and disentangle the different effects arising from regional differences regarding income distribution, namely a political effect and a tax base effect. Although they assume full income heterogeneity between individuals, they do not allow for conflict nor for partial decentralization of the redistribution policy. Le Breton and Weber (2003) examine compensation schemes that prevent a threat of secession by the regions forming a country, focusing on the trade-off between heterogeneity of preferences and economies of scale in larger countries[2]. They establish that secession-proof transfer schemes will entail some form of partial equalization among regions. Even though they do study the mechanisms through which a secession can be avoided, they do not allow for the possibility of conflict.

The analysis of Spolaore (2008a) is the most closely related to ours. He also analyzes the choice of regional conflict inputs in a context of separatist tensions, focusing on the traditional trade-off between economies of scale and heterogeneity of preferences. However, he does not address the issue of interregional inequality, nor does he allow for the possibility of an intermediate peaceful solution between unification and secession.

On the empirical side, a few attempts have been made in order to analyze the effects of decentralization on the stability of states and federations. Among them, the most related study to our analysis is the one of Bakke and Wibbels (2006), who focus on differences across federal states, and find that fiscal decentralization increases the likelihood of conflict when there are wide disparities

[2] On transfers designed so as to prevent secessions, see also Alesina and Spolaore (2003, chapter 4).

in income across regions. Also related to our argument, Tranchant (2008), although he does not focus on interregional inequality, finds that fiscal decentralization has a better conflict-mitigating impact in richer countries. Finally, Brancati (2006) finds that although decentralization reduces the probability of secession, it also has the converse indirect effect of encouraging secessionism through promoting the growth of regional political parties.

The rest of the paper is structured as follows: Section 2 describes the basic setup and analyzes the costs and benefits of seceding versus unifying in both the poor and the rich region. In Section 3, we describe the conflict technology and solve for the equilibrium of the conflictual outcome, together with its comparative statics. Section 4 introduces the possibility of using partial decentralization as a peaceful alternative to conflict, and analyses the political sustainability of such a peaceful compromise. In Section 5, we analyze the issue of self-enforceability of the peaceful outcome, while in Section 6, we discuss the implications of assuming resource constraints in the choice of conflict efforts. Finally, Section 7 concludes. Proofs and analytical derivations, when needed, can be found in the appendix.

4.2 The Choice Between Unifying and Seceding

The setup we will use to assess the regional incentives to secede or unify is a modified version of the standard model of nation formation developed by Alesina and Spolaore (1997)[3].

Suppose that the population of unit mass is located on the interval $[0, 1]$, which is also the policy line. A policy is a location of a publicly supplied public good. There are two regions of same size, and w.l.o.g., region 1 is in the interval $[0, 1/2]$, while region 2 is in the interval $[1/2, 1]$. Suppose that all individuals in region 1 are located at $1/4$, and all individuals in region 2 are located at $3/4$, which also corresponds to their ideal point regarding the location of the public good. Individual income is the same for all individuals within a region, but differs across regions, and we assume, w.l.o.g. that region 1 is the rich region (i.e. $y_1 > y_2$).

The public policy consists in a public good g whose level is fixed, with a fixed production cost k which is shared equally among individuals in a lump-sum manner. The location of g is decided upon by majority voting. Individuals value both private and public consumption, and suffer a disutility from the distance between their ideal point and the public good.

Suppose, furthermore, that under unification there is direct interregional redistribution taking place, and let the redistributive parameter be $\tau \in [0, 1]$, which we assume to be exogenous. Individual income in region $j = 1, 2$ under unification is thus given by

$$y_j^U = y_j(1-\tau) + \tau\frac{(y_1 + y_2)}{2}$$

and, clearly,

[3] See also Goyal and Staal (2004).

$$y_1 > y_1 - \frac{\tau(y_1 - y_2)}{2} = y_1^U$$

$$y_2 < y_2 + \frac{\tau(y_1 - y_2)}{2} = y_2^U$$

Let us pause here and clarify the nature of taxation and redistribution as assumed in this model. Public expenditure in private goods and services is financed by a variety of taxes. We assume that total tax collection in each region is proportional to its income, by a factor τ. The public supply of goods and services is done on an egalitarian basis so that each individual of either region obtains a benefit of $\frac{\tau(y_1+y_2)}{2}$ in the case of a centralized administration. However, under a decentralized administration, not only the location of —the now local— public good is decided at the regional level, but also a share of the publicly supplied private goods and services will be channelled through the regional administrations, with the corresponding retention of the tax collection. Throughout the paper, we shall assume that the composition of the different taxes is not the object of political debate —although the tax revenues might be retained in some proportion by the regional governments— so that the proportionality factor, τ, between tax collection and income remains unchanged[4].

Under unification, we assume that the public good g is located at 1/2, which is a natural compromise between the ideal points of the two regions under majority voting, given their equal population size. The utility of an individual i in region $j = 1, 2$ under unification is given by

$$U_j^U = y_j^U - k + g(1 - ad_i) = y_j^U - k + g(1 - \frac{a}{4})$$

where d_i is the distance between i's location and the public good, and $a \leqslant 1$ is a parameter.

Under secession, the location of g coincides with individuals' ideal point in both regions (i.e. 1/4 in region 1 and 3/4 in region 2), while the cost per capita of providing the public good is now given by $2k$, since it has to be shared among the individuals located in the region. Furthermore, there is no more redistribution between regions. The utility of an individual in region $j = 1, 2$ is thus given by

$$U_j^S = y_j - 2k + g$$

Therefore, an individual in the rich region seeks secession as long as $4k - ag < 2\tau(y_1 - y_2)$, while an individual in the poor region seeks secession as long as $ag - 4k > 2\tau(y_1 - y_2)$, so that a potential conflict over the choice of border configuration occurs whenever

$$\frac{ag}{4} - \frac{\tau(y_1 - y_2)}{2} < k < \frac{ag}{4} + \frac{\tau(y_1 - y_2)}{2} \tag{4.1}$$

[4] In Section 8.2 of the appendix, we assume proportional financing of the public good instead of direct redistribution between regions, and show that the two are equivalent.

Notice that the traditional trade-off between economies of scale and heterogeneity arises regarding the optimal choice of border configuration from the perspective of both regions. That is, the bigger the economies of scale in the production of g, and/or the smaller the disutility from distance, the more likely that they both prefer unification to secession. In other words, they have the same incentives regarding the variations of the latter parameters.

Conversely, an increase in interregional inequality, for a given level of redistribution, makes the incentives of the two regions go in opposite directions. More specifically, an increase in inequality makes unification more profitable for the poor region, while it is the opposite for the rich region. The higher inequality between regions, therefore, the more likely that there is a conflict regarding the choice between unification and secession (i.e. the more likely that (4.1) is satisfied). In particular, if inequality is high enough, the rich region seeks secession[5].

Finally, given that we assume linear utilities, the only thing that matters for determining the socially efficient border configuration is the trade-off between economies of scale and heterogeneity. In particular, unification is more efficient than secession if and only if

$$k > \frac{ag}{4} \tag{4.2}$$

That is, if the (per capita) cost of producing the public good is bigger than the average distance from it under unification, a union is efficient.

4.3 Secessionist Conflict

Assume that condition (4.1) is satisfied, so that there is a conflict of interests. We will see that condition (4.2) is in fact the key condition regarding the properties of the equilibrium with respect to interregional income inequality. To fix ideas, observe that we might have two possible situations. That is, we might have conflict under an inefficient union (i.e. when the local public good effect dominates), and so

$$\frac{ag}{4} - \frac{\tau(y_1 - y_2)}{2} < k < \frac{ag}{4}$$

or, conversely, we might have conflict under an efficient union (i.e. when the economies of scale effect dominates), and so

$$\frac{ag}{4} < k < \frac{ag}{4} + \frac{\tau(y_1 - y_2)}{2}$$

Suppose then, that the regions can invest resources into conflict so as to force the border configuration they prefer. More specifically, regions $j = 1, 2$ can choose to devote an amount F_j of

[5]Notice that in this setup, there is only one conflict configuration that can arise, namely, the rich region wants to secede whereas the poor region prefers the country to remain unified. However, this needs not be the case in alternative setups (see, for instance, Bolton and Roland (1997) for such a counter-example in the context of pure redistributive policy). We discuss this issue further in Section 5 in the context of resource constraints.

resources to conflict. As it is standard in the literature, we shall assume that the contest success function (CSF) is given by[6]

$$\pi = \frac{F_1}{F_1 + F_2}. \tag{4.3}$$

The probability of secession under conflict is increasing in the conflict effort of region 1 (i.e. F_1) and decreasing in the one of region 2 (i.e. F_2)[7].

Observe that a ratio CSF such as (4.3) is such that the winning probabilities depend on the ratio of conflict efforts F_i/F_j, $j = 1, 2$. An alternative specification is the logistic function, characterized by $\pi = \frac{e^{\beta F_1}}{e^{\beta F_1} + e^{\beta F_2}}$, where $\beta > 0$, so that the winning probabilities depend on the difference between conflict efforts $(F_i - F_j)$, $j = 1, 2$ (Hirshleifer (1989)).

One key difference between the two specifications lies in the analytical implications when only one player exerts a positive conflict effort. With a ratio function, the side making no effort faces a zero probability of winning, while this is not necessarily the case with the logistic specification. As discussed by Spolaore (2008a), given that a successful secession, when opposed by the other region, can only be obtained by active separatist effort, a ratio function appears to be more appropriate in this particular context.

Suppose there is a regional leader who chooses the aggregate regional effort invested in conflict, so as to maximize the average expected welfare of the region (or, equivalently, the expected utility of the representative individual). Formally, the leader in region $j = 1, 2$ chooses F_j so as to maximize

$$EU_j = \pi(y_j - 2k + g) - 2F_j + (1 - \pi)\left[y_j^U - k + g(1 - \frac{a}{4})\right] \tag{4.4}$$

Notice that, in this way, we abstract from the free-riding issue regarding individual contributions to conflict effort. This can be seen as if there were some kind of coordination among individuals[8].

Observe, furthermore, that we do not impose a budget constraint to individuals regarding the choice of conflict inputs F_j. In that sense, the investment in conflict should be understood as an effort, rather than as a pure monetary investment. That is, we do not assume that region 1 is more powerful than region 2 just as a result of being richer. Rather, we make the assumption that both the poor and the rich region have the same conflict capacity, and thus being poorer or richer does not influence the resources devoted to conflict in a direct manner. In that sense, conflict should be understood here in a broad manner, as it can represent any kind of individual mobilization and

[6]This class of contest success functions was first proposed by Tullock (1980) and later axiomatized by Skaperdas (1996). See Garfinkel and Skaperdas (2007) and the references therein for a description of the possible ways of modeling the conflict technology. In Section 8.3.3 and 8.3.4 of the appendix, we allow for alternative specifications for the probability of winning. In particular, we allow for asymmetric and constant CSF. All the results and intuitions of the analysis go through those alternative specifications.

[7]Spolaore (2008a) also uses this expression for the probability of secession in the same context as ours.

[8]For a discussion on individual contributions to conflict and intra-group cohesion, see Esteban and Ray (2011).

protest, rather than its somewhat restrictive interpretation as a money investment in weapons[9].

From the FOC, we get the region's best responses $F_1(F_2)$ and $F_2(F_1)$ which are given by

$$F_1(F_2) = \frac{1}{2\sqrt{2}}\sqrt{F_2(ag - 4k + 2\tau(y_1 - y_2)} - F_2$$
$$F_2(F_1) = \frac{1}{2\sqrt{2}}\sqrt{F_1(4k - ag + 2\tau(y_1 - y_2)} - F_1$$

For a given level of conflict effort in the other region, and for a given tax rate τ, more inequality translates into an increase in conflict intensity in both regions, since they both have higher stakes in such conflict. Conversely, when either a or g increases, or/and when k decreases, the rich region has more to win out of seceding, so that $F_1(F_2)$ increases, whereas the poor region has less to win out of unifying, so that $F_2(F_1)$ decreases. Observe, furthermore, that

$$\tfrac{\partial F_1(F_2)}{\partial F_2} > 0 \text{ if and only if } F_2 < \tfrac{1}{32}\left[ag - 4k + 2\tau(y_1 - y_2)\right]$$
$$\tfrac{\partial F_2(F_1)}{\partial F_1} > 0 \text{ if and only if } F_1 < \tfrac{1}{32}\left[4k - ag + 2\tau(y_1 - y_2)\right]$$

That is, from both regions' perspective, given that devoting resources to conflict is costly, it is worth increasing the effort in conflict as a best response to such an increase in the other region (i.e. conflict efforts are strategic complements) if and only if what is at stake is big enough.

Solving for the Nash equilibrium of this simultaneous game, we finally obtain the equilibrium conflict inputs F_1^* and F_2^* which are given by

$$F_1^* = \frac{\left[ag - 4k + 2\tau(y_1 - y_2)\right]^2 \left[4k + 2\tau(y_1 - y_2) - ag\right]}{128\tau^2(y_1 - y_2)^2}$$
$$F_2^* = \frac{\left[ag - 4k + 2\tau(y_1 - y_2)\right] \left[ag - 4k - 2\tau(y_1 - y_2)\right]^2}{128\tau^2(y_1 - y_2)^2}$$

In equilibrium, conflict effort is strictly positive in both regions, and which region devotes more resources to conflict depends on whether unifying is efficient or not. If a union is efficient (i.e. (4.2) holds), it follows that the poor region has bigger stakes in conflict than the rich region. Indeed, by unifying, the poor region wins both from redistribution and from the common financing of the public good, which more than compensates the loss in terms of distance. In contrast, by seceding, the rich region wins from the absence of redistribution, but it incurs a loss from not being able to exploit economies of scale in the production of the public good. As a result, when unifying is efficient, the poor region invests more in conflict than the rich one. Clearly, when seceding is the efficient solution, the opposite holds. That is, the rich region devotes more resources to conflict than the poor one.

[9] We discuss this alternative interpretation of conflict as well as the potential effects of resource constraints in section 5.

Observe that conflict effort in *both* regions is increasing in $(y_1 - y_2)$ and τ. Indeed, since more inequality increases the stakes in conflict for both the poor and the rich region, they are both willing to invest more resources in conflict. The properties of the aggregate amount of conflict in the economy is described in the following proposition:

Proposition 1 (Conflict Intensity)**.** *The total amount of resources devoted to conflict in the economy is increasing in interregional inequality. Furthermore, it is increasing in the public good and the disutility from distance, and decreasing in the cost of the public good, if and only if a union is efficient.*

The relationship between conflict intensity and the public good parameters follows the same logic. If unifying is efficient, we saw that $F_1^* < F_2^*$, that is, conflict effort is higher in the poor region. If k increases further, then, unification gets even more efficient, and thus the poor region has even more to win, while it is the opposite for the rich region. As a result, total conflict intensity decreases. Said in other words, in an efficient union, more benefits from economies of scale translate into more asymmetry in regional stakes, so that conflict is less intense. Likewise, in an efficient union, higher heterogeneity costs (i.e. higher a) translate into less asymmetry in regional stakes, so that total conflict intensity increases correspondingly.

The ratio of equilibrium conflict inputs is given by

$$\frac{F_1^*}{F_2^*} = \frac{2\tau(y_1 - y_2) + ag - 4k}{2\tau(y_1 - y_2) + 4k - ag}$$

Again, observe that no matter whether economies of scale more than compensate heterogeneity costs (and thus independently on whether the ratio is smaller or bigger than 1), an increase in inequality unambiguously makes the ratio closer to 1. That is, the higher interregional inequality, the less asymmetry in the regional stakes. As a result, regional conflict efforts get closer to each other, and the above ratio approaches 1. In the limit, when $k = \frac{ag}{4}$ (i.e. unifying is as efficient as seceding), the conflict game becomes a zero-sum game, and the two regions devote exactly the same resources to conflict (i.e. $F_1^* = F_2^*$).

Notice that, interestingly, higher conflict intensity, no matter whether it results from inequality, economies of scale or heterogeneity costs, always goes hand in hand with more uncertain associated outcomes. Suppose that unifying is efficient. In that case, the ratio F_1^*/F_2^* is strictly smaller than 1. Given that unifying is efficient, more heterogeneity costs (i.e. an increase in a) translate into more conflict intensity (Proposition 1). But then, in addition to that, F_1^*/F_2^* increases (i.e. it gets closer to 1), and thus the conflict outcome is more uncertain.

Now, observe that the same happens with inequality. We saw that, in any case, inequality always increases conflict intensity. If unifying is efficient, F_1^*/F_2^* is smaller than 1, and the ratio is increasing in inequality. That is, conflict gets both more intense and more uncertain with inequality. If, on the contrary, seceding is efficient, the ratio is correspondingly bigger than 1, it is now decreasing in inequality, and the same conclusion applies. We summarize this finding in the next proposition:

Proposition 2. *The more intense is the separatist conflict, the more uncertain is its outcome (i.e. the closer are the winning probabilities to* 1/2*).*

Finally, we can solve for the equilibrium probability of secession, which is given by

$$\pi^* = \frac{1}{2} + \frac{ag - 4k}{4\tau(y_1 - y_2)}$$

and its properties are described in the following proposition:

Proposition 3 (Equilibrium Probability of Secession)**.** *Assume k is such that there is conflict. If a union is efficient, the equilibrium probability of secession $\pi^* \in (0, 1/2)$ is increasing in inequality and redistribution. If a union is inefficient, $\pi^* \in (1/2, 1)$ is decreasing in inequality. Furthermore, no matter whether a union is efficient or not, π^* is increasing in the public good and the disutility from distance, and decreasing in the cost of the public good.*

Proof. Direct from the analytical expression of π^*. □

Quite intuitively, the probability of a successful secession occurring through conflict is always decreasing in the cost of the public good, and decreasing in the disutility from distance. If a union is efficient, the poor region invests more in conflict than the rich one, and thus $\pi^* < 1/2$. If k increases, then, the asymmetry of regional stakes gets bigger. As a result, conflict intensity decreases, and so does the ratio of conflict inputs F_1^*/F_2^* (and hence the probability of secession). If, on the contrary, it is efficient to secede, it follows than $F_1^* > F_2^*$ and hence $\pi^* > 1/2$. In that case, an increase in k reduces the asymmetry of regional stakes. As a result, conflict intensity increases, whereas the ratio F_1^*/F_2^* gets smaller, meaning that a secession is less likely.

Regarding inequality, observe that, in any case, the probability of secession always gets closer to 1/2 as inequality increases, that is, the conflict outcome is more uncertain. Whether π^* is increasing or decreasing in inequality, in turn, depends on whether unifying is efficient or not. Again, we saw that, in any case, an increase in inequality always brings the regional stakes closer to each other (and hence brings the ratio F_1^*/F_2^* closer to 1). If a union is efficient, then, more inequality brings about an increase in the latter ratio, meaning that π^* is increasing in inequality, thus approaching its upper bound 1/2 further. If, on the contrary, a secession is efficient, more inequality translates into a lower value of F_1^*/F_2^*, meaning that π^* is decreasing in inequality, thus approaching its lower bound 1/2 further.

To sum up, it turns out that it is all about the relative stakes in conflict. If the two regions have the exact same stakes in the conflict (i.e. there is no public good, for instance), they invest the same amount in such conflict, the intensity of conflict is maximized, and they both face a probability of winning of exactly 1/2. Conversely, if the regional stakes are very asymmetric, conflict intensity is sharply diminished, and one region faces a high probability of winning[10].

[10] We show in Section 8.3.2. of the appendix that the same conclusion applies regarding asymmetry in regional population size. That is, assuming that one region is bigger than the other, we obtain that conflict intensity is

We have thus showed that the symmetry of regional stakes renders conflict more intense. Interestingly, though, it turns out that more symmetric stakes actually means *more* inequality. That is, asymmetry in regional income in fact translates into symmetry in regional stakes[11]. Furthermore, while the link between inequality is clear, namely, that inequality fuels conflict in *both* the poor and the rich region, it is nevertheless ambiguous whether the probability of a successful secession is decreasing or increasing in inequality. That is, more conflict intensity can render a secession either more or less likely, while, in any case, the conflict outcome gets more uncertain.

If $k > \frac{a}{4}$, it follows directly that unification welfare-dominates the conflictual outcome, since the former is the socially desirable outcome. Then, if $k < \frac{a}{4}$ (i.e. unifying is not efficient), unification is still socially better than conflict if and only if

$$F_1^* + F_2^* > \pi^*(\frac{ag}{4} - k)$$

Observe that the more income inequality between regions, the more likely that unification welfare-dominates the conflictual outcome. This is so for two reasons. First, as we saw, inequality fuels conflict in both regions, which is totally wasteful, and second, it turns out that inequality actually makes the efficient outcome less likely under conflict. Indeed, if unifying is efficient, more inequality increases the probability of a successful secession, while it is the opposite when secession is the most efficient outcome. In other words, interregional inequality also renders the conflictual outcome less attractive socially as a result of making the socially desirable outcome less likely.

Given that interregional inequality is harmful for welfare under the possibility of conflict, the question is whether there exists an intermediate institutional arrangement between unification and secession such that both regions are prevented from initiating a conflict. In particular, can partial decentralization constitute a way to reconcile the two regions' interests, so that they both prefer to be in peace rather than starting a conflict? If the answer is yes, is such a solution politically sustainable, or, in other words, is it going to be implemented in a direct democracy? We aim at answering these questions in the next section.

4.4 Allowing for Partial Decentralization

> *Decentralization - the transfer of authority and responsibility for public functions from the central government to intermediate and local governments or quasi-independent government organizations and/or the private sector - is a complex multifaceted concept* (The Worldbank).

maximized in both regions when they are of equal population size. This is consistent with empirical findings showing that conflict is more intense in countries that are roughly equally split between two ethnic groups (see, for example, Horowitz (1985)).

[11] Again, since that means that the poorer the poor region, the more resources it devotes to conflict, notice that the assumption of no resource constraints (or, say, the interpretation of conflict as a protest rather than as a violent war) is crucial here.

Decentralization is a rather vague term which embraces a variety of concepts and typologies. One of them distinguishes between political, administrative, fiscal, and market decentralization. While political decentralization aims to give citizens or their elected representatives more power in public decision-making, administrative decentralization seeks to redistribute authority, responsibility and financial resources for providing public services among different levels of governance. Finally, fiscal decentralization relates to the dispersal of financial responsibility, and can take various forms, such as intergovernmental transfers that shift general revenues from taxes collected by the central government to local governments.

What we mean here by partial decentralization is the fact of giving more autonomy to the regions, not only regarding the policy they can implement, but also with respect to the means at their disposal in order to do it. In other words, under decentralization, a given region is left with more freedom to implement its preferred policy, while it has to rely on its own income to a bigger extent in order to do so. It follows, then, that decentralization, even though it permits to better target local preferences, also has the adverse effect of exacerbating inequality between regions. But then, given that decentralization increases inequality, which in turn creates conflict, we conjecture that its relevance as a conflict-mitigating strategy should be dependent on the underlying structure of inequality in the country.

Formally, suppose the public good can be partially decentralized, and let the degree of decentralization be $\delta \in (0,1)$. Furthermore, suppose that decentralization also applies to individual income, that is, the extent of direct interregional redistribution decreases with the degree of decentralization[12].

Let $y_j(\delta)$ be the individual income in region $j = 1, 2$ under unification with decentralization, and be given by

$$y_j(\delta) = \delta y_j + (1-\delta) y_j^U$$

Therefore, we have $y_1^U < y_1(\delta) < y_1$ and $y_2 < y_2(\delta) < y_2^U$, that is, partial decentralization impoverishes the poor region (and makes the rich region richer) by a factor δ.

The utility of an individual located in region $j = 1, 2$ under partial decentralization is given by

$$U_j(\delta) = y_j(\delta) - 2\delta k - (1-\delta)k + g\left[1 - (1-\delta)\frac{a}{4}\right]$$

Observe that $\delta = 0$ corresponds to unification, whereas $\delta = 1$ is equivalent to seceding. Partial decentralization thus constitutes an intermediate solution between those two extreme outcomes, and is introduced in order to avoid wasteful secessionist conflict. Such conflict can be avoided when individuals in both regions are better off under peace than under conflict. That is, partial decentralization can serve as a way to achieve peace if and only if both regions are prevented

[12] In fact, it can be shown that, if only applied to the location of g, decentralization cannot serve as a mean to achieve peace, the reason being that the rich region is always better off starting a conflict for any level of partial decentralization $\delta \in (0,1)$.

from initiating the conflict. We now look for the threshold levels of decentralization such that the individuals in both regions are better off under the decentralized outcome than under conflict.

An individual in region 1 prefers peace to conflict if and only if $U_1(\delta) > EU_1$, that is, if and only if

$$\delta > \left[\frac{ag - 4k + 2\tau(y_1 - y_2)}{4\tau(y_1 - y_2)}\right]^2 = \delta_1$$

Similarly, an individual in region 2 prefers peace to conflict if and only if $U_2(\delta) > EU_2$, that is, if and only if

$$\delta < \frac{[ag - 4k + 2\tau(y_1 - y_2)]\,[4k + 6\tau(y_1 - y_2) - ag]}{[4\tau(y_1 - y_2)]^2} = \delta_2$$

In order to understand the properties of the above decentralization thresholds with respect to inequality and the public good parameters, observe that utility under decentralization for an individual in region $j = 1, 2$ can be written in the following way:

$$U_j(\delta) = \delta(y_j + g - 2k) + (1 - \delta)\left[y_j^U + g(1 - \frac{a}{4}) - k\right]$$

By looking at this expression, and comparing it to (4.4), it is clear that the decentralization thresholds will have the exact same properties as the equilibrium probability of secession π^*. That is, whatever increases the odds of success of a given region in case of conflict also makes this region less willing to remain in peace. Suppose, for instance, that k increases. We saw that, no matter whether unifying is efficient or not, the probability of secession decreases correspondingly. At this bigger k then, it follows that a smaller decentralization level is required in order to prevent the poor region from starting a conflict, which means that there is less opportunity for peace in that region. Similarly, given that its odds of success are now smaller would a conflict occur, the rich region is more willing to remain in peace after the increase in k. As a result, the decentralization level above which it would start a conflict decreases. In other words, there is more room for peace in the rich region.

Suppose now that inequality increases. If a union is efficient, we saw that the probability of secession also increases. By the same reasoning, it follows that peace gets more attractive from the perspective of the poor region (i.e. δ_2 increases), while peace gets less attractive from the perspective of the rich region (i.e. δ_1 increases also). Likewise, if seceding is efficient, the probability of secession is decreasing in inequality, and so are the decentralization thresholds δ_1 and δ_2.

Finally, the difference between the two thresholds is given by

$$\delta_2 - \delta_1 = \frac{1}{2}\left\{1 - \left[\frac{ag - 4k}{2\tau(y_1 - y_2)}\right]^2\right\} > 0$$

Therefore, for any level of inequality, there is always a range of decentralization levels such that peace can be achieved, which follows directly from the fact that conflict is wasteful. By definition, any partial decentralization level $\delta \in [\delta_1, \delta_2]$ is such that $U_j(\delta) > EU_j$, $j = 1, 2$, and thus constitutes a peaceful outcome that we will call politically *sustainable*.

While the comparative statics of both decentralization thresholds δ_1 and δ_2 are the same as the ones of π^*, the properties of the *size* of the interval $[\delta_1, \delta_2]$ are directly related to the ones of total conflict intensity. More specifically, and quite intuitively, whatever makes conflict more intense also makes the interval bigger, since conflict is a pure waste of resources. That is, the more resources are wasted in conflict, the more room there is for a peaceful compromise[13].

Given that conflict gets increasingly wasteful as inequality and redistribution increase (Proposition 1), the interval $[\delta_1, \delta_2]$ increases correspondingly. Likewise, if a union is efficient, we have seen that an increase in k, by bringing about more asymmetry in regional stakes, decreases conflict intensity. As a result, given that less resources are wasted in conflict, there is also less opportunity for peace (i.e. the size of the interval $[\delta_1, \delta_2]$ decreases with k if and only if $k > \frac{a}{4}$).

The following proposition summarizes our results regarding the possibility of maintaining peace through partial decentralization conditionally on inequality and redistribution:

Proposition 4 (Decentralization and Inequality in the Shadow of Conflict)**.** *For all inequality levels, there is always a range of partial decentralization levels such that peace is politically sustainable, which is given by $\delta \in [\delta_1, \delta_2]$. Furthermore,*

1. *the minimum and maximum levels of decentralization such that peace is sustainable are decreasing in inequality if and only if a union is inefficient*
2. *the range of decentralization levels such that peace is sustainable is increasing in inequality*

Theoretically, a government could always implement a decentralization level $\delta \in [\delta_1, \delta_2]$ so as to maintain peaceful unification of the country. However, the question that naturally arises is whether such a peace-preserving decentralization level would be implemented under a direct democracy. The answer to this question is given in the following proposition:

Proposition 5. *The peaceful outcome is politically implementable in both regions, and the resulting partial decentralization level δ that is implemented depends on which region is decisive. In particular, region 1 implements δ_2 rather than starting a conflict, whereas region 2 implements δ_1 rather than starting a conflict.*

Proof. An individual in region 1 prefers peace to conflict as long as $\delta > \delta_1$. Therefore, if $\delta \in (\delta_1, \delta_2)$, an individual in region 1 does not want to start a conflict. If $\delta > \delta_2$, this is still true, but individuals

[13] In section 8.3.2 of the appendix, and consistently with this intuition, we show that when the winning probabilities are insensitive to conflict efforts (and thus $F_1^* = F_2^* = 0$), both decentralization thresholds are equal and take the same value as the constant probability π^*. Again, the fact that $\delta_1 < \delta_2$ (i.e. peace is achievable) follows directly from the fact that resources are wasted in conflict. If the latter is costless, or if the optimal strategy is to choose $F_j^* = 0$, $j = 1, 2$, it turns out that $\delta_1 = \delta_2$ and conflict cannot be avoided through partially decentralizing.

in region 2 now initiate a conflict, thus making individuals in region 1 strictly worse off. Therefore, an individual in region 1 would choose the maximum level of decentralization $\delta > \delta_1$ such that region 2 does not start a conflict, that is, $\delta = \delta_2$ (if we assume that under indifference, peace is chosen rather than conflict). By the same reasoning, an individual in region 2 would choose the minimum level of decentralization $\delta < \delta_2$ such that region 1 does not start a conflict, that is, $\delta = \delta_1$. □

Whatever assumption we make regarding who is decisive on implementing δ, it turns out that the level of partial decentralization that is implemented as a peaceful solution does depend on the level of inequality[14].

More generally, we have shown that an increase in income inequality between regions has two distinct effects: One the one hand, the range of decentralization levels such that peace is sustainable is bigger. Said in other words, the more inequality between regions, the more space there is to achieve a peaceful compromise that is sustainable. On the other hand, the level of decentralization that is implemented so as to avoid conflict can be either decreasing or increasing in inequality, depending on whether economies of scale more or less than compensate heterogeneity costs. If unifying is efficient, more inequality brings about more decentralization, whereas the opposite holds when seceding is efficient.

Proposition 6. *The level of partial decentralization that is implemented in order to prevent separatist conflict is increasing in inequality if and only if unifying is efficient.*

Proof. From Proposition 5, we know that individuals in both regions prefer to implement a threshold value of decentralization rather than starting a conflict. In particular, region 1 wants to implement δ_2, whereas region 2 wants to implement δ_1. Whichever region is decisive on that matter, therefore, or whatever compromise they reach in $[\delta_1, \delta_2]$, the result follows directly from the fact that both δ_j, $j = 1, 2$, have the same properties regarding inequality (Proposition 4). □

Proposition 4 and 6 are illustrated in Figure 4.1 and 4.2, which show the properties of the decentralization thresholds regarding inequality under an efficient and an inefficient union respectively.

Finally, total welfare under peace is higher than aggregate expected utility under conflict, that is, $W(\delta) > EW$, if and only if

$$\delta > (<) \tfrac{[ag-4k+2\tau(y_1-y_2)][3ag-12k-2\tau(y_1-y_2)]}{8\tau(ag-4k)(y_1-y_2)} = \delta_3 \text{ and } k < (>) \tfrac{ag}{4}$$

Observe that, for high inequality, no matter whether a union is efficient or not, the decentralized outcome always welfare-dominates the conflictual one for all $\delta \in (0, 1)$ (see Figure 4.1 and 4.2). This is intuitive. The higher inequality, the more resources wasted in conflict in both regions, the lower the probability of an efficient outcome, and so the more likely that the decentralized outcome yields higher welfare. As inequality keeps increasing, for given τ, the threshold leaves the interval $[0, 1]$,

[14] Given that we assumed that g is located at $1/2$ under majority voting, a natural assumption would be that the level of decentralization that is implemented in a direct democracy is $(\delta_1 + \delta_2)/2$.

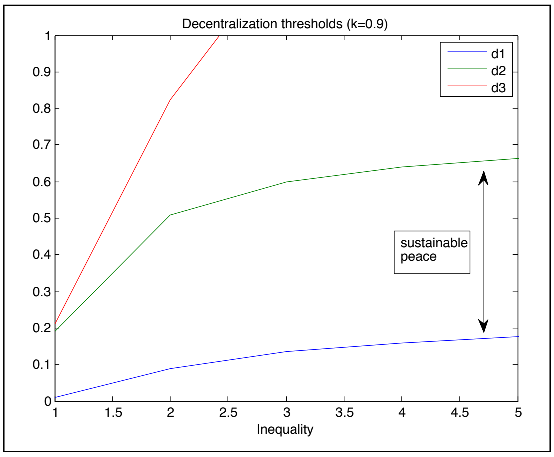

Figure 4.1: Decentralization thresholds in an efficient union

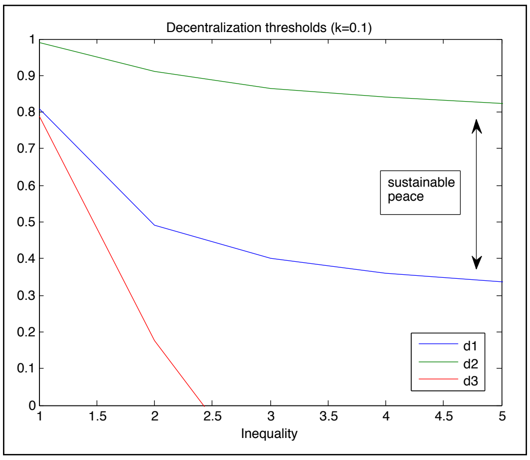

Figure 4.2: Decentralization thresholds in an inefficient union

meaning that total welfare is unambiguously higher under peace for any $\delta \in (0, 1)$. That is, the expected benefit of the best outcome occurring never compensates the loss of welfare in terms of resources lost in conflict.

Comparing the threshold levels of decentralization, we have that $\delta_3 < \delta_1 < \delta_2$ under an inefficient union and $\delta_1 < \delta_2 < \delta_3$ under an efficient union. Therefore, and quite trivially, for all the range of decentralization levels such that peace can be politically maintained, the peaceful outcome strictly welfare-dominates the conflictual outcome. More interestingly, observe that the higher inequality, the bigger the range of decentralization levels such that the best solution from a welfare perspective is not sustainable in one of the two regions.

4.5 Self-Enforcement

In the previous section, we have obtained the levels of decentralization that would prevent a secessionist conflict. We have thus implicitly assumed that the two parties —two regions or also government versus secessionists— can and do credibly commit to this agreement and hence to not reopen conflict in future. Let us take here the second interpretation, that is, one player is the central government and the other player is a potentially secessionist region. In this case, there are good reasons to assume asymmetry in the behavior of the two players from the point of view of their capacity to commit. The government can implement decentralization by a constitutional change, very hard to reverse. Also, one can think of this as one of the many agreements a government has with third parties, within the country as well as internationally. Therefore, we shall assume that the government makes concessions that are not reversible. In contrast, we shall make the extreme assumption that potentially secessionist regions do not give up on their ultimate goals and hence may not feel committed by a decentralization agreement.

We can also have such an asymmetric behavior under the interpretation that the players are the two regions: poor and rich. It could well be that the poor region does not have the political power to impose a reduction in regional autonomy in the decentralized union even in case of victory. That is, despite the fact that the poor region might have won the contest, once the country has been decentralized (and thus regional power increased correspondingly), it is likely that both regions can act as veto players on, in particular, (de)centralization matters, making decentralization an irreversible process.

In this section, we shall start by checking whether the peaceful (decentralized) outcome is self-enforcing, that is, whether there are no incentives left to start a conflict once partial decentralization has been implemented. In order to do so, we distinguish between two situations. First, we study the issue of self-enforcement assuming that full unification (i.e. $\delta = 0$) is still an available option once decentralization has been implemented. This is equivalent to assuming that neither of the two players is credibly committed to the agreement and hence that the pre-agreement scenario can be imposed in case of victory by either party. Second, we conduct the same analysis assuming this time that once the country has been partially decentralized, it is not possible to go back to a fully centralized unified country, that is, assuming that decentralization is an irreversible process. This corresponds to the case in which the government does credibly commit while the rich region does not.

So, suppose first that it is theoretically possible to go back to the unified outcome once $\delta \in [\delta_1, \delta_2]$ has been implemented. Given that $U_2^U > U_2(\delta)$ and $U_1^S > U_1(\delta)$ for all $\delta \in (0, 1)$, we have to make sure that the two regions are not willing to start a conflict under the decentralized outcome.

If full unification is still an option under decentralization, it follows trivially that peace is self-enforcing. Indeed, even though $U_2^U > U_2(\delta)$ and $U_1^S > U_1(\delta)$ for all $\delta \in [\delta_1, \delta_2]$, we know that, by definition, any such decentralization level is such that $U_j(\delta) > EU_j$, $j = 1, 2$. Since the potential conflict that would arise now is exactly the same as before (i.e. the stakes of conflict remain the same), this inequality is for sure satisfied.

Proposition 7. *If going back to unification is possible once decentralization has been implemented, peace is self-enforcing for all* $\delta \in [\delta_1, \delta_2]$.

Proof. Direct from the fact that the conflict game is exactly the same as before (i.e. EU_j, $j = 1, 2$, are left unchanged), the only difference being that the status quo is now decentralization, which is irrelevant for the choice of F_j, $j = 1, 2$. □

Suppose now that once decentralization is implemented, unification is not an option any longer. Given that $U_1^S > U_1(\delta)$ for all $\delta \in (0, 1)$, we have to make sure that the rich region is not willing to start a conflict under the decentralized outcome. As $U_2^S < U_2(\delta)$ for all $\delta \in (0, 1)$, we are again in a situation of disagreement, and thus of potential conflict. Decentralization being the (irreversible) status quo, we now ask whether the rich region is willing to start a conflict in order to force secession.

Decentralization being now the alternative to secession, individuals choose conflict efforts F_j, $j = 1, 2$ so as to maximize

$$EU_j(\delta) = \pi(y_j - 2k + g) + (1 - \pi)\Big\{y_j(\delta) - 2\delta k - (1 - \delta)k + g\left[1 - (1 - \delta)\frac{a}{4}\right]\Big\} - 2F_j$$

where, as before, $\pi = \frac{F_1}{F_1+F_2}$. Solving for the equilibrium, we get that

$$F_1^* = (1 - \delta)\frac{[ag - 4k + 2\tau(y_1 - y_2)]^2\,[4k + 2\tau(y_1 - y_2) - ag]}{128\tau^2(y_1 - y_2)^2}$$

$$F_2^* = (1 - \delta)\frac{[ag - 4k + 2\tau(y_1 - y_2)]\,[ag - 4k - 2\tau(y_1 - y_2)]^2}{128\tau^2(y_1 - y_2)^2}$$

while the equilibrium probability of secession is the same as before, and is given by

$$\pi^* = \frac{1}{2} + \frac{ag - 4k}{4\tau(y_1 - y_2)}$$

Equilibrium conflict inputs are just the same quantities as before multiplied by $(1 - \delta)$. As decentralization makes conflict less attractive for both regions (i.e. the absolute stakes are smaller), they invest strictly less in conflict. However, as the *relative* stakes are unaffected, the probability of secession remains the same.

The question is now whether it is rational for the rich region to initiate a conflict. As it turns out that $EU_1(\delta) > U_1(\delta)$ for all $\delta \in (0, 1)$, the answer to this question is positive, and hence peace is not self-enforcing.

Proposition 8. *If going back to full unification is not possible once decentralization has been implemented, peace is not self-enforcing for any* $\delta \in [\delta_1, \delta_2]$.

Proof. Direct from the fact that $EU_1(\delta) > U_1(\delta)$ for all $\delta \in (0,1)$. □

Therefore, it turns out that if full unification is not an option any longer once decentralization has been implemented, the rich region always has an incentive to fight in order to secede. In such a case, decentralization, even though it decreases the intensity of conflict, cannot serve as a way to fully eliminate it. Decentralizing decreases the absolute stakes of the two regions, but does not affect the equilibrium probability of secession should a conflict start, nor does it affect the willingness of the rich region to fight. As a result, in the absence of commitment to remain in peace conditional on decentralization, peace is not achievable through the use of partial decentralization when it is an irreversible process.

Given that the rich, secessionist region cannot credibly commit not to start a conflict for any $\delta \in [\delta_1, \delta_2]$, we would like to know in turn whether any decentralization could be chosen by majority voting. As for all $\delta \in (0,1)$, $EU_1(\delta) > EU_1$ and $EU_2(\delta) < EU_2$, it turns out that the rich region is always willing to decentralize (and start a conflict afterwards), while the poor region always opposes it. That is, the poor region, knowing that the rich will start a conflict once any δ is implemented, opposes *any* degree of decentralization in the first place, since fighting when decentralization is the status quo makes the poor region strictly worse off than if unification is the status quo. Therefore, if both regions have to agree on decentralizing, no δ is ever implemented, and conflict arises in equilibrium with unification as the status quo.

Proposition 9. *If going back to full unification is not possible once decentralization has been implemented, the poor region opposes the implementation of any* $\delta \in (0,1)$*, and separatist conflict occurs in equilibrium with unification as the status quo.*

Proof. We know from Proposition 8 that peace is not self-enforcing for any $\delta \in (0,1)$, since the rich region always starts a conflict to force secession. Then, as $EU_2(\delta) < EU_2$ for all $\delta \in (0,1)$ (i.e. the poor region is better off fighting under unification than under decentralization), the poor region opposes the implementation of any $\delta \in (0,1)$. As a result, the country is not decentralized, and assuming that (4.1) holds, the rich region starts a conflict in order to secede. □

4.6 A Note on Resource Constraints

As decentralization increases interregional inequality, a possibility is that it also makes the rich region more powerful under the assumption that the regions cannot invest in conflict more than their resources at hand. That is, if conflict takes the form of a monetary investment, we then have to take into account resource constraints when solving for the conflict equilibrium. In particular, if there are resource constraints regarding the choice of conflict inputs, the fact that $y_j(\delta) \neq y_j^U$, $j = 1, 2$, might have an effect on the enforceability of peace by means of partial decentralization.

Suppose that the two regions cannot invest in conflict more than their resources at hand. Under resource constraints, introducing decentralization might thus have an additional effect: given that decentralizing makes the rich region richer and the poor region poorer, it could influence the amount invested in conflict in a direct manner. That is, decentralization now potentially has the additional effect of making the rich region more powerful, and thus also more prone to conflict. As a result, conflict inputs now depend directly on δ, and so does the equilibrium probability of secession π^*.

If conflict takes the form of a concrete monetary investment, expected utility under conflict for an individual in region $j = 1, 2$ becomes

$$EU_j = \pi(y_j - 2k + g - 2F_j) + (1-\pi)\Big\{(y_j - 2F_j)(1-\tau) + \tau\frac{[y_1 + y_2 - 2(F_1 + F_2)]}{2} - k + g(1 - \frac{a}{4})\Big\}$$

That is, when choosing F_j^*, the individual now has to take into account that if unification is obtained through conflict, the resulting income in the opponent region is reduced as a result of monetary resources having been wasted in conflict.

The interpretation of conflict as a pure monetary investment and the effect it has on both conflict and peace enforceability (given that there are resource constraints) is a topic in itself and lies beyond the scope of the present paper. Therefore, we will just give here an intuition of what could possibly happen, and leave the full analysis for future research.

So, suppose, as before, that the status quo is unification, and that the rich region is willing to start a conflict so as to force secession. Suppose, then, that some decentralization level $\delta \in (\delta_1, \delta_2)$ compatible with peace is implemented, so that the available income of the poor region is now $y_2(\delta) < y_2^U$. That is, decentralization makes the poor region poorer. In order for peace to be self-enforcing, it must be the case that none of the two regions is willing to start a conflict under the decentralized outcome.

A possibility, then, is that the optimal choice of conflict investment F_2^* that maximizes EU_2 as defined above is now bigger than the available resources of the poor region under decentralization. As a result, the poor region invests all its resources in conflict. Formally, we have

$$F_2^* = \tfrac{y_2(\delta)}{2}$$

so that the best response of the rich region is just

$$F_1^* = F_1(y_2(\delta))$$

Therefore, the possibility now arises that decentralization, rather than being an *alternative* to conflict, might well affect conflict in a direct manner[15]. For instance, if the poor region is constrained as a result of decentralization, and under the assumption that war is a rational option for the rich region, the resources devoted to conflict in the poor region will be a decreasing function of δ.

[15] That is, contrary to the case where investment in conflict is an effort, the conflict equilibrium (F_1^*, F_2^*, π^*) would now depend on δ.

As a result, the investment of the rich region will also depend directly on δ, and a conjecture is that the resulting probability of secession will be increasing in the decentralization parameter (i.e. decentralization makes the rich more powerful)[16].

In such a case, the question in turn would be whether the $\delta \in (\delta_1, \delta_2)$ implemented under a direct democracy is such that the poor region remains unconstrained, in which case we could conclude that peace is self-enforcing. Observe, furthermore, that δ_2 is likely to be much bigger than before, since part of the income of the rich region is wasted in war under forced unification, implying that redistribution is clearly less beneficial for the poor region when there is a conflict, which makes it a less attractive option.

Although we are being informal and rather speculative in the reasoning above, our aim is to point out an additional potential effect of using partial decentralization as a mean to mitigate separatist conflict: if conflict is considered as a monetary investment, then decentralization, by exacerbating interregional inequality, also makes the rich region more powerful (i.e. its odds of success get higher), and thus more prone to conflict[17]. This means that decentralization might actually influences conflict in a direct manner, and this in turn clearly has an effect on the willingness of the two regions to remain in peace. In particular, partial decentralization might not be such a good way of achieving peace in such a case.

In fact, Spolaore (2008b) has already pointed out this possible adverse effect of partial decentralization on separatist conflict. However, in his model, conflict investment is exogenous and it is assumed, also in an exogenous fashion, that decentralizing always makes the *seceding* region (the 'periphery") more powerful. We believe, however, that this is not necessarily the case. Strictly speaking, and assuming that one - or even two - regions are constrained, decentralizing makes the *rich* region more powerful. Whether the region seeking secession is the rich one in fact depends on the model at hand, as well as on the parameters' constellation. For instance, Bolton and Roland (1997), considering only pure income redistribution between individuals, have shown that the poor region might prefer seceding to unifying, even though it does not receive transfers from the rich one. Therefore, to the extent that the region seeking secession might in fact be the poor one, more decentralization might well be associated with *less* conflict and a *lower* probability of secession in equilibrium.

4.7 Concluding Remarks

We have analyzed the possibility of using partial decentralization as a (secessionist) conflict-mitigating strategy in the presence of interregional inequality. According to our results, while

[16]Notice that there are many other possibilities. For instance, it might be the case that the poor region is already constrained under unification, so that $F_2^* = \frac{y_2^U}{2}$. Alternatively, the rich region might also be constrained under decentralization, and thus $F_j^* = \frac{y_j(\delta)}{2}$ for both $j = 1, 2$, and so on.

[17]Therefore, decentralizing would have three distinct effects on conflict: first, it brings the public good closer to regional tastes (both regions are less prone to conflict). Second, it exacerbates interregional inequality (the poor region is more prone to conflict). Third, it makes the rich region more powerful (the rich region is more prone to conflict).

implementing some degree of partial decentralization can indeed prevent the rich region from initiating a separatist conflict, it should not be too high, since the resulting impoverishment of the poor region in turn makes it willing to start the conflict.

We showed that the properties of the range of partial decentralization levels such that peace is sustainable do depend on the underlying structure of inequality in the country. In particular, more inequality between regions is associated with, on the one hand, more room for peace in equilibrium, and, on the other hand, with either a lower or a higher degree of partial decentralization that is implemented so as to avoid conflict. More specifically, being in an (in)efficient union implies a (negative) positive relationship between inequality and the degree of decentralization in the shadow of separatist conflict.

We showed that there is always a range of decentralization levels that prevents separatist conflict for all levels of inequality in the unified country. However, this turns out not to be the case if institutional rules are such that full unification is not an option any longer once the country has been partially decentralized. In such a case, given that the rich region would always start a conflict under decentralization, it follows that the poor region prevents any decentralization from being implemented in the first place, so that conflict occurs in equilibrium with unification as the status quo. Hence, the possibility of using partial decentralization as a conflict-mitigating strategy depends crucially on whether decentralizing is an institutionally and politically reversible process.

One implication of our results is that empirical studies which aim at exploring the relationship between decentralization and conflict should condition the nature of this relationship on the underlying structure of inequality between regions. Bakke and Wibbels (2006) have already made an attempt in this direction. However, since according to our results, the assertion according to which the combination of high interregional inequality and a high degree of decentralization contributes to conflict, is verified only in the case of an inefficient union, we believe that one should also control for the composition of public expenditures. That is, if economies of scale in the public sector are important, it might well be that both high decentralization and high interregional inequality are in fact compatible with peace. Conversely, if the heterogeneity costs associated to the uniformity of policies in a union are important, decentralization should rather be limited in the presence of wide income disparities across regions.

We believe there are several directions in which our model could be developed further. First, we assumed that conflict effort is coordinated, implying that individual contributions to conflict are equal within a region. This assumption is clearly questionable, especially if we add heterogeneity with respect to individual location inside a region[18]. That is, one can see as rather unlikely that all individuals in a region will accept to devote the same amount of resources to conflict, given that the associated expected benefits vary among individuals within the region. Most likely, under

[18]In Section 8.3.5 of the appendix, we extend the model to allow for such heterogeneity. In particular, we show that under coordinated conflict effort, heterogeneous location of individuals within a region causes a majority in the poor region to prefer full decentralization to conflict at low levels of inequality. Furthermore, the size of $[\delta_1, \delta_2]$ is decreasing in inequality.

any kind of intra-regional heterogeneity, the issue of free-riding regarding individual contributions is an important one, so that one should solve the model using individual best responses rather than assuming a coordinated conflict effort.

Second, we have assumed equal individual income within a region. Alternatively, as in Bolton and Roland (1997), we could introduce intra-regional income heterogeneity, meaning that there would be both losers and winners within each region stemming from a national redistribution policy under unification. That is, in addition to interregional transfers of resources, there would be transfers from rich to poor individuals in each region. Furthermore, preferences for redistribution would then differ across regions. As a result, the costs and benefits of seceding versus unifying would be affected in several manners, and so would be the incentives of the two regions to start a conflict[19] .

Third, while we have considered a nation formed of two regions in our setup, the model could be generalized to allow for three or even for n regions (as in Panizza (1999)), opening the possibility of alliance formation and coalitions between regions who could then fight to secede jointly.

Fourth, we have assumed a static framework. As suggested by Spolaore (2008a), one could imagine more complex dynamic games, in which a region may influence future decisions of the other region by pre-committing to a given level of conflict effort. For example, the poor region could invest in conflict at an earlier stage in order to prevent the rich one from investing in separatist activities at a further stage.

Finally, while we have discussed the possibility of conflict as a monetary investment and the potential role of resource constraints, we did not analyze those issues analytically. However, as secession attempts often take the form of violent civil wars, we believe that investigating the possibility of achieving peace through partial decentralization when conflict intensity is directly related to regional wealth is a necessary step to complete our analysis. In particular, in the presence of resource constraints, the degree of partial decentralization and the intensity (and outcome) of conflict would then be linked in a direct fashion.

4.8 References

1. Alesina, A. and E. Spolaore (1997), On the Number and Size of Nations, *Quarterly Journal of Economics*, Vol. 112, Issue 4, pp. 1027-1056.

2. Alesina, A. and E. Spolaore (2003), *The Size of Nations*, The MIT Press.

3. Bakke, K and E. Wibbels (2006), Diversity, Disparity, and Civil Conflict in Federal States, *World Politics*, Vol. 59, No. 1, pp. 1-50.

4. Bolton, P. and G. Roland (1997), The Breakup of Nations: A Political Economy Analysis, *Quarterly Journal of Economics*, Vol. 112, No. 4, pp. 1057-1090.

[19] For an attempt to explore empirically the link between inequality *within* regions and the demand for sovereignty, see Sambanis and Milanovic (2009).

5. Brancati, D. (2006), Decentralization: Fueling the Fire or Dampening the Flames of Ethnic Conflict and Secessionism?, *International Organization*, Vol. 60, Issue 3, pp. 651-685.

6. Esteban, J. and D. Ray (2011), Linking Conflict to Inequality and Polarization, *American Economic Review*, Vol. 101, Issue 4, pp. 1345-74.

7. Garfinkel, M.R. and S. Skaperdas (2007), Economics of Conflict: An Overview, in T. Sandler and K. Hartley (Eds.), *Handbook of Defense Economics*, Vol. II, pp. 649-709.

8. Goyal, S. and K. Staal (2004), The Political Economy of Regionalism, *European Economic Review*, Vol. 48, Issue 3, pp. 563-593.

9. Grossman, H.I. and M. Kim (1995), Swords or Plowshares? A Theory of the Security of Claims to Property, *Journal of Political Economy*, Vol. 103, Issue 6, pp. 1275-288.

10. Gurr, T.R. (2000), Peoples versus States: Minorities at Risk in the New Century, US Institute of Peace Press, Washington DC.

11. Hirshleifer, J. (1989), Conflict and Rent-Seeking Success Functions: Ratio versus Difference Models of Relative Success, *Public Choice*, Vol. 63, pp. 101-112.

12. Horowitz, D. L. (1985), *Ethnic groups in conflict*, University of California Press, Berkeley.

13. Lake, D.A. and Rotchild, D. (2005), Territorial Decentralization and Civil War Settlements, *Sustainable Peace: Power and Democracy after Civil Wars*, edited by Philip G. Roeder and Donald Rothchild, Ithaca, NY, Cornell University Press.

14. Lazear, E..P., and S. Rosen (1981), Rank-Order Tournaments as Optimum Labor Contracts, Journal of Political Economy, Vol. 89, No. 5, pp. 841-864.

15. Le Breton, M. and S. Weber (2003), The Art of Making Everybody Happy: How to Prevent a Secession?, *IMF Staff Papers*, Vol. 50, No. 3, pp. 403-435.

16. Linz, J.J. and A. Stepan (2000), Inequality Inducing and Inequality Reducing Federalism, Paper presented at the World Congress of the International Political Science Association, Quebec City, Quebec, Canada.

17. Myerson, R. (1981), Optimal auction design, *Mathematics of Operations Research*, Vol. 6, No. 1, pp. 58-73.

18. Panizza, U. (1999), On the Determinants of Fiscal Centralization: Theory and Evidence, *Journal of Public Economics*, Vol. 74, Issue 1, pp. 97-139.

19. Sacchi, A. and S. Salotti (2011), Income inequality, Regional Disparities, and Fiscal Decentralization in Industrialized Countries, paper prepared for the *Workshop on Regional and Urban Economics: Inequality and Regional Growth and Cohesion*, University of Barcelona.

20. Sambanis, N. and B. Milanovic (2009), Explaining the Demand for Sovereignty, Manuscript, Department of Political Science, Yale University.

21. Skaperdas, S. (1996), Contest Success Functions, *Economic Theory*, Vol. 7, No. 2, pp. 283-90.

22. Spolaore, E. (2008a), Civil Conflict and Secessions, *Economics of Governance*, Vol. 9, No. 1, pp. 45-63.

23. Spolaore, E. (2008b), Federalism, Regional Redistribution, and Country Stability, *Studies in Fiscal Federalism and State-Local Finance*, Edward Elgar Publishing.

24. Tranchant, J-P. (2008), Fiscal Decentralization, Institutional Quality, and Ethnic Conflict: A Panel Data Analysis, 1985-2001, *Conflict, Security and Development*, Vol.8, No. 4, pp. 419-514.

25. Tullock, G. (1980), Efficient Rent-Seeking, in Buchanan, J.M., Tollison, R.D. and G. Tullock (Eds.), *Towards a Theory of a Rent-Seeking Society*, Texas A&M University Press, pp. 97-112.

4.9 Appendix

4.9.1 Proofs

Proof of Proposition 1. The individuals in region 1 choose F_1 so as to maximize their expected utility, that is, so as to maximize

$$EU_1 = \pi(y_1 - 2k + g) - 2F_1 + (1-\pi)\left[y_1^U - k + g(1-\frac{a}{4})\right]$$

$$\Leftrightarrow EU_1 = \frac{F_1}{(F_1+F_2)}(y_1 - 2k + g) - 2F_1 + \frac{F_2}{(F_1+F_2)}\left[y_1 - \tau\frac{(y_1-y_2)}{2} - k + g(1-\frac{a}{4})\right]$$

Similarly, the individuals in region 2 choose F_2 so as to maximize their expected utility, that is, so as to maximize

$$EU_2 = \pi(y_2 - 2k + g) - 2F_2 + (1-\pi)\left[y_2^U - k + g(1-\frac{a}{4})\right]$$

$$\Leftrightarrow EU_2 = \frac{F_1}{(F_1+F_2)}(y_2 - 2k + g) - 2F_2 + \frac{F_2}{(F_1+F_2)}\left[y_2 + \tau\frac{(y_1-y_2)}{2} - k + g(1-\frac{a}{4})\right]$$

Taking derivatives, we obtain

$$\frac{\partial EU_1}{\partial F_1} = \frac{F_2}{(F_1+F_2)^2}(y_1 - 2k + g) - 2 - \frac{F_2}{(F_1+F_2)^2}\left[y_1^U + g(1-\frac{a}{4}) - k\right]$$

$$\frac{\partial EU_2}{\partial F_2} = -\frac{F_1}{(F_1+F_2)^2}(y_2 - 2k + g) - 2 + \frac{F_1}{(F_1+F_2)^2}\left[y_2^U + g(1-\frac{a}{4}) - k\right]$$

From the FOC, we get the regions' best responses, which are given by

$$F_1(F_2) = \frac{1}{2\sqrt{2}}\sqrt{F_2(ag - 4k + 2\tau(y_1 - y_2)} - F_2$$
$$F_2(F_1) = \frac{1}{2\sqrt{2}}\sqrt{F_1(4k - ag + 2\tau(y_1 - y_2)} - F_1$$

and, substituting into one another, we obtain the equilibrium conflict inputs:

$$F_1^* = \frac{[ag - 4k + 2\tau(y_1 - y_2)]^2\,[4k + 2\tau(y_1 - y_2) - ag]}{128\tau^2(y_1 - y_2)^2}$$
$$F_2^* = \frac{[ag - 4k + 2\tau(y_1 - y_2)]\,[ag - 4k - 2\tau(y_1 - y_2)]^2}{128\tau^2(y_1 - y_2)^2}$$

The aggregate amount of conflict in the economy is thus given by

$$F_1^* + F_2^* = \frac{1}{8}\left[\tau(y_1 - y_2) - \frac{(ag - 4k)^2}{4\tau(y_1 - y_2)}\right]$$

Taking derivatives with respect to $(y_1 - y_2)$ and τ, we obtain

$$\frac{\partial(F_1^* + F_2^*)}{\partial(y_1 - y_2)} = \frac{1}{8}\left[\tau + \frac{(ag - 4k)^2}{4\tau(y_1 - y_2)^2}\right] > 0$$

$$\frac{\partial(F_1^* + F_2^*)}{\partial\tau} = \frac{1}{8}\left[(y_1 - y_2) + \frac{(ag - 4k)^2}{4\tau^2(y_1 - y_2)}\right] > 0$$

$$\frac{\partial(F_1^*+F_2^*)}{\partial a} = \frac{g(4k-ag)}{16\tau(y_1-y_2)} > 0 \text{ if and only if } k > \frac{ag}{4}$$

$$\frac{\partial(F_1^*+F_2^*)}{\partial g} = \frac{a(4k-ag)}{16\tau(y_1-y_2)} > 0 \text{ if and only if } k > \frac{ag}{4}$$

$$\frac{\partial(F_1^*+F_2^*)}{\partial k} = \frac{ag-4k}{4\tau(y_1-y_2)} > 0 \text{ if and only if } k < \frac{ag}{4}$$

□

Proof of Proposition 4. The decentralization thresholds in region 1 and 2 are respectively given by

$$\delta_1 = \left[\frac{ag - 4k + 2\tau(y_1 - y_2)}{4\tau(y_1 - y_2)}\right]^2$$

$$\delta_2 = \frac{[ag - 4k + 2\tau(y_1 - y_2)]\,[4k + 6\tau(y_1 - y_2) - ag]}{[4\tau(y_1 - y_2)]^2}$$

Taking derivatives, we obtain the following expressions:

$$\frac{\partial\delta_1}{\partial(y_1-y_2)} = \frac{(4k-ag)[ag-4k+2\tau(y_1-y_2)]}{8\tau^2(y_1-y_2)^3} > 0 \text{ if and only if } k > \frac{ag}{4}$$

$$\frac{\partial \delta_1}{\partial \tau} = \frac{(4k-ag)[ag-4k+2\tau(y_1-y_2)]}{8\tau^3(y_1-y_2)^2} > 0 \text{ if and only if } k > \frac{ag}{4}$$

and, for the second threshold,

$$\frac{\partial \delta_2}{\partial (y_1-y_2)} = \frac{(ag-4k)[ag-4k-2\tau(y_1-y_2)]}{8\tau^2(y_1-y_2)^3} < 0 \text{ if and only if } k > \frac{ag}{4}$$

$$\frac{\partial \delta_2}{\partial \tau} = \frac{(ag-4k)[ag-4k-2\tau(y_1-y_2)]}{8\tau^3(y_1-y_2)^2} < 0 \text{ if and only if } k > \frac{ag}{4}$$

Finally, the difference between the two thresholds is given by

$$\delta_2 - \delta_1 = \frac{1}{2}\left\{1 - \left[\frac{ag-4k}{2\tau(y_1-y_2)}\right]^2\right\} > 0$$

which is clearly increasing in $(y_1 - y_2)$ and τ.

□

Proof of Proposition 7. Equilibrium expected utility under conflict for an individual in region 1, when going back to unification is not possible, is given by

$$EU_1(\delta) = \pi^*(y_j - 2k + g) - 2F_1^* + (1-\pi^*)\left\{y_j(\delta) - 2\delta k - (1-\delta)k + g\left[1 - (1-\delta)\frac{a}{4}\right]\right\}$$

while his utility under the decentralized outcome is given by

$$U_1(\delta) = y_1(\delta) - 2\delta k - (1-\delta)k + g\left[1 - (1-\delta)\frac{a}{4}\right]$$

Substituting for equilibrium values (F_1^*, F_2^*, π^*) of the conflict game, and taking the difference between utilities, we get that

$$EU_1(\delta) - U_1(\delta) = \frac{(1-\delta)\left[ag - 4k + 2\tau(y_1-y_2)\right]^3}{\left[8\tau(y_1-y_2)\right]^2} > 0$$

Therefore, $EU_1(\delta) > U_1(\delta)$ for all $\delta \in (0,1)$, meaning that for any level of decentralization, the rich region always starts a conflict. That is, peace is not self-enforcing for any $\delta \in (0,1)$ when going back to unification is not possible once decentralization has been implemented.

□

4.9.2 Proportional Financing of the Public Good

Suppose now that the public good is financed with a proportional tax on individual income. As a result, and given that the two regions are of equal population size, the rich region, even though it benefits from economies of scale under a union, also has to pay a bigger share of the total cost of producing g (i.e. more than half of k). This means that, in addition to the direct redistribution taking place through the tax rate τ, there is also indirect redistribution taking place through the proportional financing of g.

For tractability, however, and in order to isolate this latter effect, we assume here that there is no direct redistribution between regions under unification (i.e. $\tau = 0$). Under the assumption of proportional financing of g, the utility of an individual in region $j = 1, 2$ under secession and unification is now given by

$$U_j^U = y_j \left[1 - \frac{2k}{(y_1 + y_2)} \right] + g(1 - \frac{a}{4})$$

$$U_j^S = y_j(1 - \frac{2k}{y_j}) + g = y_j - 2k + g$$

The condition for a conflict to occur (i.e. such that the rich region seeks secession while the poor region seeks unification) is now given by[20]

$$\frac{ag(y_1 + y_2)}{8y_1} < k < \frac{ag(y_1 + y_2)}{8y_2} \tag{4.5}$$

while, as before, a union is efficient whenever

$$k > \frac{ag}{4} \tag{4.6}$$

If (7) holds, then the equilibrium probability of secession when a conflict occurs is given by

$$\pi^* = \frac{ag(y_1 + y_2)}{8k(y_1 - y_2)} - \frac{y_2}{y_1 - y_2}$$

and so, as in the case of lump sum taxation, $\pi^* < 1/2$ and is increasing in inequality if and only if a union is efficient (i.e. (9) holds).

Therefore, it turns out that interregional inequality, being through direct or indirect redistribution, has the same effect on the conflict equilibrium. If the tax used to finance g is proportional, more inequality induces more indirect redistribution, the absolute stakes of the two regions are bigger, and thus both regions increase their conflict effort. Then, as more inequality also has the effect of bringing the relative stakes of the two regions closer to each other, it also brings the probability of a successful secession closer to 1/2 (which implies an increase or a decrease in π^* depending on whether a union is efficient or not).

Given that there are economies of scale in the production of g together with proportional taxation, the utility of an individual in region $j = 1, 2$ under partial decentralization is now given by

$$U_j(\delta) = y_j \left[1 - \delta\frac{2k}{y_j} - (1 - \delta)\frac{2k}{(y_1 + y_2)} \right] + g \left[1 - (1 - \delta)\frac{a}{4} \right]$$

[20] Observe that by assuming different region size here, the possibility could arise that it is the *poor* region that seeks secession. Clearly, this would have many implications for the analysis. In particular, and as already mentioned in section 5.2, assuming resource constraints in that case could imply that decentralizing makes the *unifying* region more powerful. We leave the analysis of these issues for future research.

Notice that, even though decentralization does not apply directly to individual income any longer, it still impoverishes the poor region in an indirect manner, since the latter benefits from the common - more than half - financing of g to a lesser extent.

Again, the fact that conflict is wasteful implies that there is always room for a peaceful compromise, so that $\delta_2(k) > \delta_1(k)$. Furthermore, the fact that conflict gets increasingly wasteful with inequality implies in turn that $(\delta_2(k) - \delta_1(k))$ is always increasing in inequality, no matter whether unifying is efficient or not. Finally, and as in the case of lump sum taxation, the decentralization thresholds $\delta_j(k)$, $j = 1, 2$, are increasing in inequality if and only if a union is efficient.

Again, it turns out that the effect of interregional inequality both on conflict and on the peaceful outcome is exactly the same no matter whether it goes through direct or indirect redistribution. If both types of redistribution are present, then, a conjecture is that the effect of inequality on both the conflictual and peaceful outcomes would be qualitatively the same as the ones just analyzed separately, but quantitatively stronger[21].

Finally, it is worth noting that direct redistribution is actually not necessary in order to generate separatist conflict driven by interregional inequality. That is, implicit redistribution taking place through proportional taxes in order to finance a common public good under unification might be sufficient to induce the rich region to start a conflict so as to force secession.

4.9.3 Extensions

In this section, we will extend the model in several directions, assuming for simplicity that there are no economies of scale in the production of the public good. Observe that this is equivalent to assuming that unifying is not efficient (i.e. $k < \frac{ag}{4}$). Since the effects of whether economies of scale more than compensate heterogeneity costs or not are now well understood, we will focus exclusively on the case of an inefficient union, which makes the analysis shorter and analytically simpler.

The Benchmark Case with no Economies of Scale

Assume that the per capita cost of producing the public good is equal to k under both unification and secession, that is, there are no economies of scale in the production of g. The utility an individual in region $j = 1, 2$ under unification is now given by

$$U_j^U = y_j^U - k + g(1 - \frac{a}{4})$$

and under secession, it is given by

$$U_j^S = y_j - k + g$$

[21] Notice, however, that direct redistribution also has the effect of decreasing the extent of indirect redistribution (in the limit, when $\tau = 1$, there is no more indirect redistribution through the financing of g under unification), so that things might in fact be more complex than it seems at first. However, the algebra involved makes the analysis intractable and does not allow us to analytically disentangle all those effects.

Observe that an individual in the rich region is always better off under secession, whereas an individual in the poor region faces a trade-off regarding the choice between secession and unification. More specifically, an individual in region 2 is better off under unification if and only if τ is high enough, that is, if and only if

$$\tau > \frac{ag}{2(y_1 - y_2)} \tag{4.7}$$

Secessionist Conflict

Suppose condition (7) is satisfied, and thus the redistributive tax rate is high and such that the rich region seeks secession, while the poor region prefers the country to remain unified. As before, the probability of secession occurring through conflict is given by

$$\pi = \frac{F_1}{F_1 + F_2}$$

Solving for the conflict equilibrium, we get the equilibrium conflict inputs, which are given by

$$F_1^* = \frac{[ag + 2\tau(y_1 - y_2)]^2 \, [2\tau(y_1 - y_2) - ag]}{128\tau^2(y_1 - y_2)^2}$$

$$F_2^* = \frac{[ag + 2\tau(y_1 - y_2)] \, [ag - 2\tau(y_1 - y_2)]^2}{128\tau^2(y_1 - y_2)^2}$$

and, finally, we can solve for the equilibrium probability of secession, which is given by

$$\pi^* = \frac{1}{2} + \frac{ag}{4\tau(y_1 - y_2)}$$

As there are no economies of scale in the production of g, it follows that a union is never efficient in this economy. As a result, the stakes in conflict are strictly higher for the rich region than for the poor one, which implies that $F_1^* > F_2^*$ and hence $\pi^* > 1/2$.

As before, the more inequality between regions, the more there is to win from fighting in both regions, and hence both F_1^* and F_2^* are increasing in inequality. That is, inequality fuels conflict in both regions. Furthermore, an increase in inequality, by decreasing the asymmetry in regional stakes, also brings the ratio F_1^*/F_2^* closer to 1. This means in turn that the probability of secession is decreasing in inequality. Therefore, and as mentioned before, assuming no economies of scale is equivalent to assuming that there are such economies and $k < \frac{ag}{4}$, since both imply that unifying is inefficient.

Partial Decentralization

The utility of an individual located in region $j = 1, 2$ is now given by

$$U_j(\delta) = y_j(\delta) - k + g \left[1 - (1 - \delta)\frac{a}{4}\right]$$

Since unifying is always inefficient, we find without surprise that both thresholds δ_1 and δ_2 are decreasing in inequality and redistribution, and increasing in the disutility from distance. Finally, the difference between the two thresholds is given by

$$\delta_2 - \delta_1 = \frac{1}{2}\left\{1 - \left[\frac{ag}{2\tau(y_1 - y_2)}\right]^2\right\} > 0$$

which means, as before, that there is always a range of decentralization levels such that peace can be achieved for any level of inequality. This range is given by the interval $[\delta_1, \delta_2]$, and we can see directly from its analytical expression that it is increasing in $(y_1 - y_2)$ and τ, and decreasing in a and g.

Therefore, in an inefficient union, an increase in income inequality between regions has two effects: On the one hand, peace is sustainable at strictly lower levels of decentralization, but, on the other hand, the range of decentralization levels such that peace is sustainable is bigger. In any case, the decentralization level that is implemented so as to prevent conflict is decreasing in inequality.

Different Region Size

Suppose now that a proportion α of individuals live in the rich region (i.e. region 1), while a proportion $(1-\alpha)$ of individuals live in the poor region (i.e. region 2), where $\alpha \in (0, 1)$. The average income in the economy is thus given by $\alpha y_1 + (1-\alpha)y_2$ and, as before, we assume that individuals in region 1 are all located at 1/4, while individuals in region 2 are all located at 3/4. Under secession, the public good is located at 1/4 and 3/4 in the rich and the poor region respectively. Now, under unification, we assume that a compromise is reached such that the public good is located at $\alpha\frac{1}{4} + (1-\alpha)\frac{3}{4}$. That is, the bigger the region in terms of its relative population, the bigger its influence on the public good location[22].

As before, an individual from the rich region is strictly better off under secession than under unification, given that he looses utility by unifying both in terms of income and in terms of distance from the public good. Conversely, an individual from the poor region prefers unification to secession if and only if

$$\tau > \frac{ag}{2(y_1 - y_2)}$$

which is exactly the same condition as before. Indeed, the smaller the poor region (i.e. the bigger α), the more each individual wins out of redistribution, whereas the more each individual looses

[22]We could alternatively assume that, as one region is bigger, it is also decisive regarding the choice of the location of g. In that case, whether it is the rich or the poor region that is bigger clearly is a key element regarding the costs and benefits of seceding versus unifying, the condition for a conflict to occur would be modified accordingly, and so would be the decentralization thresholds. However, as mentioned before, our interpretation of g is quite general, and we do not allow for complete decentralization of g under unification. That is, we assume that some centralization is always necessary to run the country under unification (the presence of federal institutions for instance). Therefore, even though having one region bigger implies that the central government is biased towards that region, the influence of the smaller region, in proportion to its size, still remains.

out of distance from g by unfying. As those two effects exactly cancel each other, the condition on whether unification is preferred to secession is the same as before. That is, α constitutes a scale factor regarding the costs and benefits of secession, but does not determine which one dominates the other.

Secessionist Conflict

In the rich region, F_1 is chosen so as to maximize

$$EW_1 = \pi(y_1 - k + g) - \frac{F_1}{\alpha} + (1-\pi)\left\{y_1(1-\tau) + \tau\left[\alpha y_1 + (1-\alpha)y_2\right] + g\left[1 - a(\alpha\tfrac{1}{4} + (1-\alpha)\tfrac{3}{4} - \tfrac{1}{4})\right] - k\right\}$$

and similarly, F_2 is chosen in the poor region so as to maximize

$$EW_2 = \pi(y_2 - k + g) - \frac{F_2}{(1-\alpha)} + (1-\pi)\left\{y_2(1-\tau) + \tau\left[\alpha y_1 + (1-\alpha)y_2\right] + g\left[1 - a(\tfrac{3}{4} - \alpha\tfrac{1}{4} - (1-\alpha)\tfrac{3}{4})\right] - k\right\}$$

Equilibrium conflict inputs are now given by

$$F_1^* = \alpha(1-\alpha)\frac{\left[ag + 2\tau(y_1 - y_2)\right]^2\left[ag - 2\tau(y_1 - y_2)\right]}{32\tau^2(y_1 - y_2)^2}$$

$$F_2^* = \alpha(1-\alpha)\frac{\left[ag + 2\tau(y_1 - y_2)\right]\left[ag - 2\tau(y_1 - y_2)\right]^2}{32\tau^2(y_1 - y_2)^2}$$

so that the equilibrium probability of secession remains unchanged and is still given by

$$\pi^* = \frac{1}{2} + \frac{ag}{4\tau(y_1 - y_2)}$$

Therefore, the only effect of allowing for different regional population size under the conflictual outcome is that the total amount of resources devoted to conflict decreases. Notice, in particular, that the conflict inputs are maximized for $\alpha = \frac{1}{2}$. Therefore, while inequality in terms of income increases conflict, inequality in terms of population size has the effect of decreasing conflict. However, the *ratio* of equilibrium conflict inputs, and thus the probability of secession, remain the same no matter the value of α, given that the relative (aggregate) stakes are not affected.

Therefore, it turns out that the more asymmetries between the two regions, the less intense is conflict. That is, more asymmetry in relative stakes (less inequality) and more asymmetry in size both reduce the intensity of conflict in the nation.

Partial Decentralization

Allowing for partial decentralization for a given α, we get the following peace-compatible decentralization thresholds for the individuals in the rich and the poor region respectively:

$$\delta > \pi^* - \frac{2F_1^*}{\alpha(1-\alpha)\left[ag + 2\tau(y_1 - y_2)\right]} = \delta_1$$

$$\delta < \pi^* - \frac{2F_2^*}{\alpha(1-\alpha)\left[2\tau(y_1-y_2)-ag\right]} = \delta_2$$

Finally, the threshold level of decentralization regarding total welfare is given by

$$\delta_3 = \pi^* - \frac{F_1^* + F_2^*}{\alpha(1-\alpha)ag}$$

Therefore, all the decentralization thresholds are the same as before, so that all previous conclusions extend to the case of different regional population size.

Fixed Probability of Winning

Suppose now that whatever the resources devoted to conflict, the probability of a successful secession is constant and given by $\pi^* = \beta \in (0,1)$. That is, the probability of winning for both regions is independent of conflict efforts. In that case, it is clear that none of the two regions is willing to invest in conflict (i.e. $F_1^* = F_2^* = 0$).

Intuitively, given that no resources are wasted in conflict, achieving peace through partial decentralization gets clearly harder. In fact, as we describe in the following proposition, there is no possibility for such peace:

Proposition 10. *If $\pi^* = \beta \in (0,1)$ for all F_j, $j = 1,2$, peace is sustainable in the poor (rich) region for all inequality levels if and only if $\delta < \beta$ ($> \beta$).*

In words, when the probability of winning is insensitive to conflict efforts, sustainable peace is not any longer achievable through partial decentralization[23]. Notice, furthermore, that total welfare is higher under peace than under conflict for all $\delta > \beta$. However, for all such decentralization levels, the poor region prefers to start a war.

Notice that the result above is conditional on individuals being risk neutral. As a result, the decentralization thresholds coincide with the probability of secession (i.e. $\delta_1 = \delta_2 = \beta$). If $\delta < \beta$, it means that decentralization gives less utility than what can be expected under conflict, and thus individuals in the rich region prefer conflict to peace. Under risk aversion, however, those individuals might prefer peace even though $\delta < \beta$, since the outcome of conflict is uncertain. In other words, if individuals are risk averse, we expect that $\delta_1 \neq \delta_2 \neq \beta$ and $\beta \in (\delta_1, \delta_2)$.

Notice, finally, that $\beta = 1/2$ corresponds to a particular case of a general symmetric contest success function of the form

$$\pi = \frac{F_1^\gamma}{F_1^\gamma + F_2^\gamma}$$

[23] Assuming that peace is chosen under indifference, peace would be sustainable only at $\delta = \beta$.

where $\gamma \in [0, 1]$, so that expected payoffs are concave in own effort[24]. Thus, $\gamma = 1$ corresponds to the CSF postulated before, while $\gamma = 0$ corresponds to $\pi = 1/2$. We do not treat the case of $\gamma \in (0, 1)$, but we are able to capture the main idea through the two limit cases, that is, the less sensitive is π to conflict efforts, the less regions invest in conflict, and thus the less room there is for peace, meaning that the size of $[\delta_1, \delta_2]$ is increasing in γ.

Asymmetric CSF

Suppose that one of the two regions has an advantage regarding the effectiveness of its conflict effort, so that the probability of secession is given by

$$\pi = \frac{\lambda F_1}{\lambda F_1 + (1 - \lambda) F_2}$$

where $\lambda \in (0, 1)$. Thus, the rich region has an advantage over the poor region whenever $\lambda > 1/2$, while the opposite holds whenever $\lambda < 1/2$ ($\lambda = 1/2$ corresponds to the symmetric case analyzed before). Observe that the incentives for seceding versus unifying in the two regions are unchanged, since only π is modified.

The Rich Region has an Advantage

Suppose that $\lambda > 1/2$, so that the rich region has an advantage in military proficiency. This could capture the fact that in many cases, war proficiency is positively correlated with wealth.

Solving for the conflict equilibrium, it turns out that both $F_j^*(\lambda)$, $j = 1, 2$, are decreasing in the military proficiency of the rich region, while the equilibrium probability of secession is now given by

$$\pi^*(\lambda) = \frac{\lambda \left[ag + 2\tau(y_1 - y_2)\right]}{2\tau(y_1 - y_2) - (1 - 2\lambda)ag}$$

and is increasing in the parameter λ. Quite intuitively, it turns out that the properties of the equilibrium regarding inequality and redistribution (as well as regarding a and g) are the same as before. In other words, introducing an advantage in military proficiency in the rich region does not change the results obtained before (i.e. when $\lambda = 1/2$). However, it translates into lower conflict in both regions, and into a bigger probability of secession should a conflict occur.

The same comment applies once we introduce partial decentralization given that $\lambda > 1/2$. Peace occurs in the interval $[\delta_1(\lambda), \delta_2(\lambda)]$, which, as before, is increasing in inequality, while both thresholds are decreasing in inequality.

The military advantage of the rich region, however, has the effect of making both decentralization thresholds strictly higher than before for a given level of inequality. That is, high decentralization levels are easier to implement, since the poor region is less prone to conflict than before. Similarly,

[24] This is the CSF proposed by Tullock (1980) and used extensively in the literature.

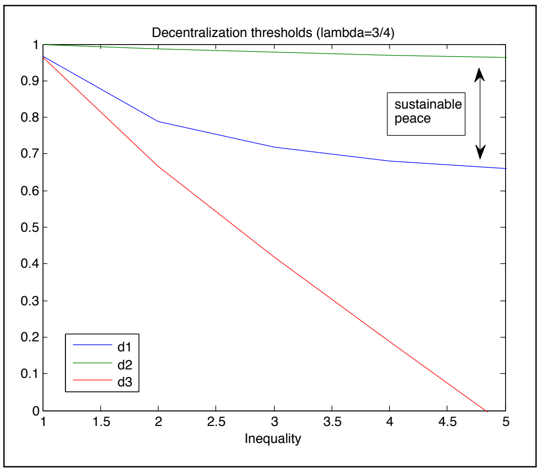

Figure 4.3: The rich region has an advantage

for high values of λ, the rich region is willing to stay in peace only at decentralization levels very close to 1. Furthermore, the size of the interval $[\delta_1(\lambda), \delta_2(\lambda)]$ is decreasing in λ, meaning that the bigger the advantage of the rich region, the less room there is for a peaceful compromise. Intuitively, this is so because both conflict inputs are decreasing in λ (i.e. conflict is less wasteful), so that a bigger λ is translates into less room for peace. Finally, notice that since conflict is less harmful than before (because there is less waste of resources together with a higher probability of secession), $\delta_3(\lambda)$ has increased for a given level of inequality.

The Poor Region has an Advantage

Suppose now that $\lambda < 1/2$, so that the conflict effort of the poor region is more efficient than the one of the rich region. To the extent that the poor region is in a defensive position vis-a-vis the rich one (recall that the status quo is unification), this could be the case if one believes that the defender has the advantage (Grossman and Kim (1995)).

Even though $\pi^*(\lambda)$ is still increasing in λ, the comparative statics of $F_j^*(\lambda)$, $j = 1, 2$ with respect to λ are now ambiguous. More importantly, it is not necessarily true that F_1^* is increasing in inequality when $\lambda < 1/2$. As it increases the stakes in conflict, more inequality for sure fuels conflict in the poor region given that its effort is relatively more efficient than the one of the rich region. As a result, then, more inequality has an ambiguous effect on the conflict effort in the rich region: there is more to win out of fighting, but the conflict effort is not so efficient, so that the total effect of inequality on F_1^* is ambiguous. In any case, however, and as before, the probability of secession is clearly decreasing in inequality.

Given that the link between inequality and conflict in the rich region is now unclear, we expect the one between peace and inequality to be modified as well. Introducing partial decentralization for $\lambda < 1/2$ and looking at the decentralization thresholds, it turns out that, as before, the three thresholds $(\delta_1(\lambda), \delta_2(\lambda), \delta_3(\lambda))$ are decreasing in inequality and redistribution. However, as it is unclear whether inequality increases the total amount of conflict, it is not necessarily true any longer that the interval $[\delta_1(\lambda), \delta_2(\lambda)]$ is increasing in inequality. That is, when the poor region has an advantage regarding the efficiency of its conflict effort, it might well be the case that more inequality makes peace harder to achieve.

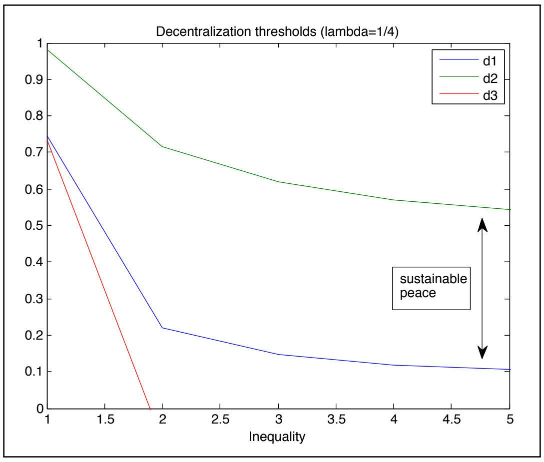

Figure 4.4: The poor region has an advantage

Finally, as in the case of $\lambda > 1/2$, both decentralization thresholds $\delta_j(\lambda)$, $j = 1, 2$, are increasing in λ. Indeed, given that π^* is increasing in λ, a bigger λ makes peace more (less) sustainable in the poor (rich) region. However, it is now ambiguous whether the interval $[\delta_1(\lambda), \delta_2(\lambda)]$ is increasing in λ or not. Since it is unclear whether λ makes conflict more or less intense, the variations of the size of the interval with respect to λ are also ambiguous. For the same reason, it is also unclear whether δ_3 is increasing in λ (a bigger λ makes a secession more likely but it might either increase or decrease the total amount of wasteful conflict).

Heterogeneity within Regions

Suppose that individuals are uniformly distributed on $[0, 1]$, and that their location coincides with their ideal point. Under secession, the regional median individuals are decisive regarding the location of g (and thus the location is $1/4$ and $3/4$), whereas under unification, the national median individual is decisive (and thus g is located at $1/2$).

As we assume that conflict effort is coordinated, we require that, for a conflict to start, individuals within a region must unanimously agree on the border configuration they prefer[25]. As shown below, the condition for a conflict to start is then the same as before, that is, $\tau > \frac{ag}{2(y_1 - y_2)}$, so that the rich region seeks secession, while the poor region seeks unification.

[25] As all individuals contribute the same amount to the regional conflict effort, it seems to us a minimum requirement that they all agree to start a conflict, even though they do not have the same benefits associated to the outcome.

Secessionist Conflict

Suppose that the last condition is satisfied. Given that the two regions disagree on whether the country should remain unified or not, as before, they can choose to invest resources in conflict so as to obtain the border configuration they prefer. However, given that individuals are now distributed uniformly on $[0, 1]$, they do not all benefit in the same way from a given border configuration.

The probability of a secession occurring is given by

$$\pi = \frac{F_1}{F_1 + F_2}$$

where F_j, $j = 1, 2$, is the total amount of conflict effort in region j. Suppose there is a group leader that chooses the regional amount of conflict effort so as to maximize the region's average expected payoff, that is, he chooses F_j so as to maximize

$$EW_j = \pi \left[y_j - k + g(1 - \frac{a}{8}) \right] - 2F_j + (1 - \pi) \left[y_j^U - k + g(1 - \frac{a}{4}) \right]$$

Solving for the conflict equilibrium, we obtain that the equilibrium probability of secession is now given by

$$\pi^* = \frac{1}{2} + \frac{ag}{8\tau(y_1 - y_2)}$$

and is thus smaller than before, since secession now has some cost in terms of average distance from the public good. Heterogeneity within regions makes the relative stakes of the 2 regions closer to each other, the ratio F_1^*/F_2^* is closer to 1, and thus secession is correspondingly less likely.

Partial Decentralization

As before, we would like to know whether wasteful separatist conflict can be avoided using partial decentralization. We will show that it is indeed the case, although the range of decentralization levels such that peace is sustainable is now decreasing in inequality. In that sense, heterogeneity makes peace harder to achieve through partial decentralization, which, as we will explain below, is due to the fact that conflict effort is coordinated within regions, whereas the expected benefits from conflict vary among its individuals.

Utility under peace for individual i in region $j = 1, 2$ is now given by

$$U_{ji}^D = y_j(\delta) - k + g \left[1 - (1 - \delta) a d_i^U - \delta a d_i^S \right]$$

where d_i^U and d_i^S are the distance between i's ideal point and the public good under unification and secession respectively. Therefore, the decentralization threshold for a given individual now depends on his distance from the public good under the different institutional arrangements. Under the conflictual outcome, individual i in region j gets expected utility

$$EU_{ji} = \pi^* \left[y_j - k + g(1 - ad_i^S)\right] - 2F_j^* + (1 - \pi^*) \left[y_j^U - k + g(1 - ad_i^U)\right]$$

The threshold levels of decentralization such that individual i in region 1 and 2 is indifferent between peace and conflict are now given by

$$\delta_{1i} = \pi^* + \frac{4F_1^*}{2a(d_i^S - d_i^U)g - \tau(y_1 - y_2)}$$

$$\delta_{2i} = \pi^* + \frac{4F_2^*}{2a(d_i^S - d_i^U)g + \tau(y_1 - y_2)}$$

Consider the individuals in region 1 (i.e. the rich region) for whom $d_i^U > d_i^S$. Those individuals are the ones located in the interval $[0, 3/8)$, and since they face no trade-off regarding the choice between seceding and unifying, their utility under peace is maximized for $\delta = 1$. Therefore, those individuals are better off under peace than under conflict as long as $\delta > \delta_{1i}$, $i \in [0, 3/8)$, where, as before, δ_{1i} is decreasing in inequality and redistribution, and increasing in the public good and the disutility from distance. Observe, furthermore, that δ_{1i} is now increasing in $(d_i^U - d_i^S)$, that is, the bigger the relative benefit of seceding in terms of distance, the bigger the decentralization level above which those individuals are willing to remain in peace.

Consider the individuals located in the interval $[0, 1/4]$. For all of them, $(d_i^U - d_i^S) = 1/4$, so that we have

$$\delta_{1i} = \pi^* - \tfrac{8F_1^*}{ag + 2\tau(y_1 - y_2)} \text{ for } i \in [0, \tfrac{1}{4}]$$

Consider now the individuals located in the interval $(1/4, 3/8)$. For those individuals, it is also true that $d_i^U > d_i^S$, but now we have that $\frac{\partial(d_i^U - d_i^S)}{\partial i} < 0$, that is, the closer is i to $1/2$, the smaller the relative benefit of seceding in terms of distance. At the limit, for $i = 3/8$, $d_i^U = d_i^S$. Therefore, we have that

$$\delta_{1,0} = ... = \delta_{1,\frac{1}{4}} > ... > \delta_{1,\frac{5}{16}} > ... > \delta_{1,\frac{3}{8}} = \pi^* - \frac{4F_1^*}{\tau(y_1 - y_2)}$$

Given that conflict effort is coordinated, so that everyone invests the same amount in conflict, a given individual is more willing to remain in peace the smaller his relative benefit from seceding in terms of his own distance from the public good. Said in other words, the bigger the individual relative benefit from seceding in terms of distance, the bigger his decentralization threshold δ_{1i} above which he's willing to stay in peace.

Consider now the individuals located in the interval $(3/8, 1/2)$. For those individuals, $d_i^U < d_i^S$, that is, they are further away from the public good under secession than under unification. It is therefore unclear whether those individuals are better off under secession than under unification in the first place. Formally, $i \in (3/8, 1/2)$ prefers secession to unification if and only if $U_{1i}^S > U_{1i}^U$, that is, if and only if

$$\frac{\tau(y_1 - y_2)}{2} > ag(d_i^S - d_i^U)$$

Since we have assumed that $\tau > \frac{ag}{2(y_1 - y_2)}$, and as $(d_i^S - d_i^U)$ is at most equal to 1/4 for all $i \in (3/8, 1/2)$, it follows directly that the above inequality is satisfied. Therefore, all those individuals also prefer secession to unification, and their utility under peace is maximized for $\delta = 1$. It follows that $i \in (3/8, 1/2)$ prefers peace to conflict as long as $\delta > \delta_{1i}$. Observe, furthermore, that $\frac{\partial(d_i^S - d_i^U)}{\partial i} > 0$ for $i \in (3/8, 1/2)$ so that $\frac{\partial \delta_{1i}}{\partial i} < 0$. Finally, since the individuals in $(3/8, 1/2)$ clearly have smaller decentralization thresholds than the ones for whom $d_i^U > d_i^S$ (i.e. $i \in [0, 3/8)$), it follows that

$$\delta_{1,0} = ... = \delta_{1,\frac{1}{4}} > ... > \delta_{1,\frac{1}{2}}$$

Therefore, we have the following result[26].

Proposition 11. *If $\delta \geqslant \delta_{1,\frac{1}{4}}$, all individuals in the rich region are better off under peace than under conflict, while if $\delta < \delta_{1,\frac{1}{4}}$, a majority of individuals start a conflict rather than staying in peace.*

Consider now the individuals in region 2 (i.e. the poor region) for whom $d_i^U < d_i^S$. Those are the ones located in the interval $[1/2, 5/8)$, and they do not face any trade-off regarding the choice between unification and secession, so that their utility under peace is maximized for $\delta = 0$. Any such individual is willing to stay in peace as long as $\delta < \delta_{2i}$. Observe, furthermore, that $\frac{\partial(d_i^S - d_i^U)}{\partial i} < 0$ for $i \in [1/2, 5/8)$, so that $\frac{\partial \delta_{2i}}{\partial i} > 0$.

For the individuals located in the interval $(5/8, 1]$, $d_i^U > d_i^S$. Such individuals prefer unification to secession as long as

$$\frac{\tau(y_1 - y_2)}{2} > ag(d_i^U - d_i^S)$$

Since we have assumed that $\tau > \frac{ag}{2(y_1 - y_2)}$, and as $(d_i^U - d_i^S)$ is at most equal to 1/4 for all $i \in (5/8, 1]$, it follows directly that the above inequality is satisfied. Therefore, all those individuals also prefer unification to secession, and their utility under peace is maximized for $\delta = 0$. It follows that $i \in (5/8, 1]$ prefers peace to conflict as long as $\delta < \delta_{2i}$. Observe that $\frac{\partial(d_i^U - d_i^S)}{\partial i} > 0$ for $i \in (5/8, 3/4)$, so that $\frac{\partial \delta_{2i}}{\partial i} > 0$ for those individuals. Then, for all $i \in [3/4, 1]$, $(d_i^U - d_i^S) = 1/4$, and thus they have a common decentralization threshold, which is strictly bigger than δ_{2i} for all $i \in [1/2, 3/4)$. Therefore, we have the following distribution of decentralization thresholds in the poor region:

$$\delta_{2,\frac{1}{2}} < ... < \delta_{2,\frac{3}{4}} = ... = \delta_{2,1}$$

Proposition 12. *If $\delta \leqslant \delta_{2,\frac{3}{4}}$, a majority of individuals in the poor region are better off under peace than under conflict, while if $\delta > \delta_{2,\frac{3}{4}}$, all individuals start a conflict rather than staying in peace.*

[26] We assume that in case of indifference, peace is chosen over conflict.

The question is now whether there exists a range of decentralization levels such that peace is sustainable in both regions, and if so, whether it would be implemented under a direct democracy in a national referendum. In order to answer those two questions, we have to order the decentralization thresholds in the country. In fact, given that $\delta_{2,\frac{1}{2}} > \delta_{1,0}$, it follows directly that

$$\delta_{1,\frac{1}{2}} < ... < \delta_{1,\frac{1}{4}} = ... = \delta_{1,0} < \delta_{2,\frac{1}{2}} < ... < \delta_{2,\frac{3}{4}} = ... = \delta_{2,1}$$

Therefore, we have the following result regarding the peace-compatible decentralization levels:

Proposition 13. *If $\delta_{1,\frac{1}{4}} \leqslant \delta \leqslant \delta_{2,\frac{1}{2}}$, peace is unanimously preferred to conflict in both regions, while if $\delta_{1,\frac{1}{4}} \leqslant \delta \leqslant \delta_{2,\frac{3}{4}}$, peace is preferred to conflict by all individuals in the rich region, and a majority in the poor region. Otherwise (i.e. $\delta < \delta_{1,\frac{1}{4}}$ or $\delta > \delta_{2,\frac{3}{4}}$), a majority in one region prefers to start a conflict rather than staying in peace.*

We assume that for decentralization to be implemented as a peaceful solution, a majority in both regions must be willing to remain in peace. The peace-compatible range of decentralization levels is thus defined by $\delta \in [\delta_{1,\frac{1}{4}}, \delta_{2,\frac{3}{4}}]$. As before, given that conflict is costly, there always exists a range of δ such that peace can be nationally maintained.

Proposition 14. *The peaceful outcome consisting in unification with partial decentralization is politically sustainable in both regions as long as $\delta \in [\delta_{1,\frac{1}{4}}, \delta_{2,\frac{3}{4}}]$. Furthermore, both thresholds, as well as the size of the interval, are decreasing in inequality and redistribution.*

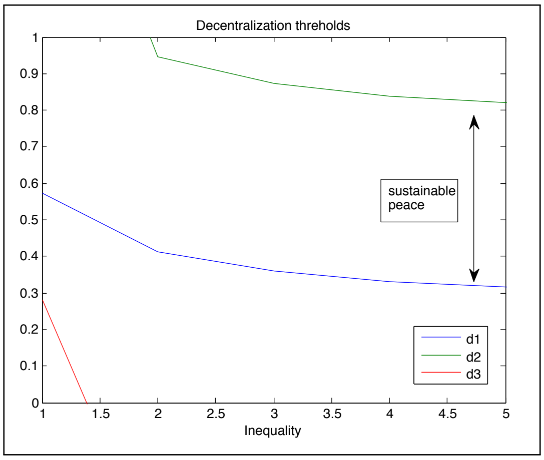

Figure 4.5: Heterogeneity within regions

As in the case of homogeneous location, since the probability of secession is decreasing in inequality, it follows that both thresholds are also decreasing in inequality. Notice, however, that for low levels of inequality, $\delta_{2,\frac{3}{4}}$ is bigger than 1, meaning that a majority in the poor region actually prefers a peaceful secession to a conflict. Recall that the decentralization threshold in the poor region is the one of the individuals located in $[3/4, 1]$. Given that those individuals are the ones benefiting the least from unifying in terms of distance, their contribution to conflict is high relative to their expected individual payoff, so that they are not so prone to conflict. In particular, for low levels of inequality (and thus few benefits in terms of redistribution), they are willing to remain in peace at full decentralization (i.e. $\delta = 1$).

More striking is the fact that the interval $[\delta_{1,\frac{1}{4}}, \delta_{2,\frac{3}{4}}]$ is now decreasing in inequality. As before, more inequality fuels conflict in both regions, and so both thresholds decrease with inequality. The difference, however, is that more inequality now also brings individual expected payoffs within a region closer to each other. That is, the individuals in $[0, 1/4]$ 'free-ride" less on average - common - contribution (and thus are less prone to conflict), whereas individuals in $[3/4, 1]$ suffer less from'free-riding" on their contribution (and thus are more prone to conflict). As a result, $\delta_{2,\frac{3}{4}}$ decreases faster than $\delta_{1,\frac{1}{4}}$, so that the size of the interval decreases with inequality.

Finally, the fact that δ_3 has decreased means that conflict under heterogeneity is more harmful than under homogeneity. From an aggregate welfare perspective, even though secession still dominates unification, it is less worthy given that it also has a cost in terms of distance. Furthermore, secession becomes less likely should a conflict occur. As a result, peace is more valuable than in the case of homogeneity, and thus δ_3 is strictly smaller.

In order to determine whether the peaceful solution is implementable politically, we ask which would be the choice of δ in a nationwide vote. As before, all individuals in the rich region choose $\delta_{2,\frac{3}{4}}$ rather than starting a conflict. That is, they choose the highest δ compatible with peace for a majority in the poor region. Similarly, all individuals in the poor region choose $\delta_{1,\frac{1}{4}}$ rather than starting a conflict. The decentralization level that is implemented thus depends on which region is decisive, and, in any case, is decreasing in inequality.