Economics
The Open-Access, Open-Assessment E-Journal

Discussion Paper
No. 2013-53 | October 07, 2013 | http://www.economics-ejournal.org/economics/discussionpapers/2013-53

Guilt Aversion and Redistributive Politics: A Moral Intuitionist Approach

Gilles Le Garrec

Abstract
In mainstream economics individuals are supposed to be driven only by their self-interest. By contrast, surveys clearly show that people do care about fairness in their demand for redistribution. In this article, in the spirit of the "new synthesis" in moral psychology (Haidt, 2007: The new synthesis in moral psychology) the author proposes to modelize the voting behavior over redistribution as the interaction between (a) an automatic cognitive process which quickly generates intuitions on the fair level of redistribution, (b) a rational self-oriented reasoning which controls the feeling of guilt associated with fair intuitions. In addition, considering that guilt aversion depends on the cultural context, the author shows that the model exhibits a multiplicity of history-dependent steady states which may account for the huge difference of redistribution observed between Europe and the United States.

Published in Special Issue Economics, Psychology and Choice Theory

JEL D03 D64 D72 H53
Keywords Redistribution; voting behavior; fairness; behavioral economics

Authors
Gilles Le Garrec, ✉ OFCE, Center for Economic Research – Sciences Po, 69 quai d Orsay, 75340 Paris, gilles.legarrec@sciencespo.fr

Citation Gilles Le Garrec (2013). Guilt Aversion and Redistributive Politics: A Moral Intuitionist Approach. Economics Discussion Papers, No 2013-53, Kiel Institute for the World Economy. http://www.economics-ejournal.org/economics/discussionpapers/2013-53

Received September 5, 2013 **Accepted as Economics Discussion Paper** September 19, 2013 **Published** October 7, 2013

1 Introduction

In mainstream economics individuals are supposed to be driven only by their self interest. As a consequence, when studying the redistributive phenomenon in democracy, the first challenge for most economists is to explain why there is so little redistribution in democracy. Indeed, considering that the idea of democracy is captured by the majority rule, as the median citizen is characterized by an income lower than the average, a majority should support a complete income redistribution to satisfy their self interest. As a canonical answer to this issue, Meltzer and Richard (1981) have shown that selfish people have no interest to support a overly high redistribution, even if they are poorer than average, because of a tax disincentive effect that lowers productivity. Their model also implies that we should observe a positive correlation between redistribution and income inequality. However, such a correlation is weakly supported by data. While redistribution is significantly higher and more progressive in (continental) Europe than in the United States, their pre-tax income inequality appear similar (see Table 1). By contrast, Perotti (1996), Moene and Wallerstein (2001), de Mello and Tiongson (2006) and Iversen and Soskice (2006) support that the empirical relationship between income inequality and redistribution is the opposite of the predicted one or is insignificant.

Countries	Pre-tax income inequality (GINI)	Public Social Spendings (% GDP)	Progressivity Index
Sweden	0.37	19.8	1.77
France	0.41	18.3	1.73
Germany	0.43	15.5	1.95
UK	0.41	15.2	1.12
US	0.43	10.6	1.31

Table 1. Income inequality (working age population) and social spending (except Old Age) in 2005 (source: OECD and author's calculation; the progressivity index gives the decrease in percent of income inequality allowed by a social spending of 1% of GDP).

In order to improve the canonical model's predictions, different dimensions have been investigated[1]. From a behavioral perspective, the postulate that individuals are driven only by their self interest has been challenged when studying redistribution (Piketty, 1995, Alesina and Angeletos, 2005, Lind, 2007). It has been challenged firstly because an impressive number of experimental studies have pointed out that individuals do not behave selfishly in the way supposed in mainstream economics (see Batson, 1991, Fehr and Schmidt, 2006). It has also been challenged because analysis of survey data clearly show that people do care about fairness in their demand for redistribution (Fong, 2001, Corneo and Grüner, 2002, Alesina and La Ferrara, 2005, Corneo and Fong, 2008, Alesina and Giuliano, 2010). In line with such findings, Alesina, Glaeser and Sacerdote (2001) therefore show that beliefs according to which *luck rather than effort determines income*[2] are strong

[1] Economic (Bénabou, 2000), political (Roemer 1998, Rodriguez, 2004, Iversen and Soskice, 2006, Petrova, 2008, Campante, 2010) and behavioral (Bénabou and Ok, 2001, Bénabou and Tirole, 2006); see Alesina and Angeletos (2004), Lind (2005) and Campante (2010) for overviews.

[2] From World Values Survey data, they highlight that 54% of Europeans versus 30% of

predictors, unlike income inequality, on the national level of redistribution. As a matter of fact, fairness has a major influence in shaping redistributive politics.

If voters care only about the welfare of the population when considering redistribution, Piketty (1995) has shown that international differences in the level of redistribution (when countries share identical economic fundamentals) can be explained by different beliefs about social mobility sustained by an imperfect learning process. Close to the concept of reciprocal altruism, Lind (2007) considers that voters care about their self interest, and the welfare of the members of their own group, more than the rest of the population. In such a context, he shows that both fractionalization and group antagonism reduce redistribution. In the spirit of Gilens (1999), he then supports that the difference of redistribution between Europe and the United States is sustained by a difference in ethnic fractionalization. In Alesina and Angeletos (2005), voters also care about both their self interest and fairness. In their model, fairness is not defined according to a utilitarian social welfare as in Piketty (1995) and Lind (2007), but according to a deontological principle (*everyone should receive what he deserves*) whose relevance is empirically supported in psychology and sociology when considering income distribution (see Schokkaert, 1998, Forsé and Parodi, 2006). With income depending on both effort and luck, they show that cultural variability of the level of redistribution arises as a multiplicity of equilibria resulting from different self-fulfilled beliefs. By expecting low redistribution, Americans invest in their human capital and generate conditions for low redistribution by reducing the importance of luck in the income determination. Conversely, by expecting a high redistribution, Europeans invest less in their human capital and will support a high redistribution later.

In this article, we consider that voters do care both about fairness (in

Americans believe that *luck rather than effort determines income*.

the way of Alesina and Angeletos, 2005) and their self interest. In addition, connected to recent findings in neurosciences and cognitive psychology, we also explicitly consider that fair and selfish motives are generated by two distinct cognitive processes, named following Kahneman (2003) *system 1* and *system 2* (see also Camerer et al., 2005, Evans, 2008), whose different features (see Table 2) are of interest in explaining cultural variability of redistributive politics. First, as argued by Haidt (2001, 2007, 2008; see also Nado et al., 2006), an automatic and domain-specific cognitive process[3] quickly generates and relates moral intuitions (shaped according to Hauser, 2006, and Mikhail, 2007, by a Universal Moral Grammar - UMG) then emotions such as guilt[4]. Emotions are then integrated into a self-oriented rational cognitive process (*system 2*; see Table 2) such as human behavior is equally driven by emotions and by the standard economic reasoning (Gray, 2004, Bechara, 2004, Cohen, 2005, Pessoa, 2008). Thereafter, considering in the cultural trend of psychology that guilt aversion is context dependent, we show that the huge difference of redistribution observed between Europe and the United States can be sustained by a multiplicity of steady states.

[3] As explained by Fehr and Schmidt (2006), "the term automatic in this case refers to a process that does not require conscious and effortfull processing but which can nevertheless be inhibited or controlled."

[4] Supporting this thesis, a large number of both neuroimaging and neuropsychological studies (Greene et al., 2001, 2004, Berthoz et al., 2002, 2006, Moll et al., 2002, 2005, Decety and Chaminade, 2003, Hsu et al., 2008; Damasio, 1994, Anderson et al., 1999, Blair, 2001, Koenigs and Tranel, 2007; see Greene, 2005, Young and Koenigs, 2007 for reviews) show that fair behaviors are associated with brain areas involved in emotional processing (amygdala, insula, ...).

system 1	**system 2**
Affective	Analytic
Automatic	Controlled
Fast	Slow
Domain specific	Domain general
Contextualized	Abstract
Evolutionary rationality	Individual rationality

Table 2. Main attributes associated with dual-process decision-making.

The rest of the paper is organized as follows. In section 2, we present basic decisions such as effort, and the *system 1* as an automatic and domain-specific cognitive process which quickly generates moral intuitions over the fair level of redistribution. In section 3, in connection with the dual-process theory of decision-making and the "new synthesis" in moral psychology (Haidt, 2007), we specify as described in Figure 1 how the feeling of guilt associated with the moral intuition is integrated into the rational reasoning when voting over redistribution. We then present conditions for the existence of a multiplicity of steady states in order to explain the huge difference of redistribution observed between Europe and the United States. In section 4, we generalize the multiplicity of steady states found in the previous section by considering family background and the intergenerational transmission of inequality. We conclude briefly in the last section.

2 Moral intuition and the fair level of redistribution

In social sciences, following Kant in philosophy and Kohlberg in psychology, morality is usually associated with a controlled, emotionless and logical reasoning. By contrast, according to Haidt (2007) moral intuition refers

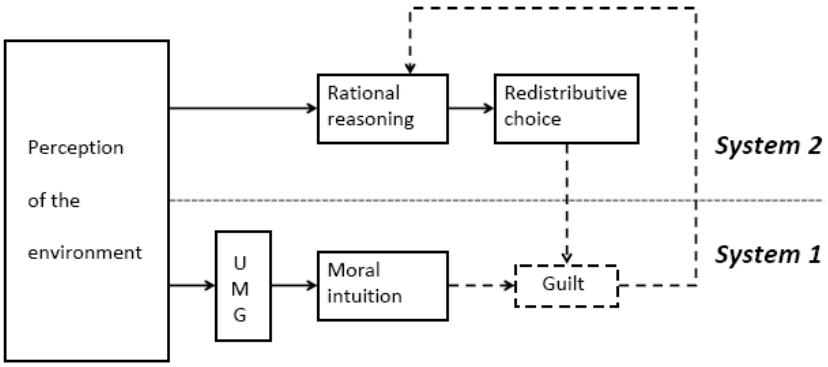

Figure 1: Dual-process decision-making and voting over redistribution

to "fast, automatic and (usually) affect-laden processes in which an evaluative feeling of good-bad or like-dislike appears in consciousness without any awareness of having gone through steps of search, weighing evidence, or infering a conclusion". In the "new synthesis" in moral psychology he describes, moral reasoning is then mostly a "post-hoc process in which we search for evidence to support our initial intuitive reaction".

2.1 An insight into the lab

As an illustration of moral intuition and the emotional processing associated with it when considering redistribution, let us consider the Ultimatum Game (Güth et al, 1982). In this game, two anonymous subjects must agree in the split of a given amount of money ($10). One participant, the proposer, can make one proposal on how to split the money. The other one, the recipient, can either accept or reject the proposal. If he accepts, the proposal is carried out, if not both players get nothing. In theory, in such a one-shot game the proposer should offer to the recipient an amount of money as low as possible, and the recipient should accept any proposal superior to 0. By contrast, robust findings in the lab show an average proposal of $4 with a large number of 50/50 splits. Besides, if all proposed 50/50 splits are

accepted, any proposal below $2.50 has a high probability to be rejected. The Ultimatum Game (with other games such as the Dictator Game and the Trust Game) stresses behaviors which are characterized by fairness and inequity aversion[5] (Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000).

In addition, Sanfey et al. (2003) have shown that the rejection of an unfair proposal (assimilated to a fair punishment) was related to the activation of the anterior insula. The rejection of an unfair proposal has also been associated with an increased skin conductance (van 't Wout et al., 2006). These findings elicit the emotional side of this recipient's choice. According to the "new synthesis" in moral psychology, an equal share of the money in the Ultimatum Game appears then as a moral intuition. However, the initial distribution of the money in the game can not account for the market income determination in real life. Indeed, in the game the proposer is initially rich only by luck. By contrast, in real life, an individual can have a high income because he works hard. Therefore, as long as personal effort changes the perception of the fair income, it follows that effort as well as luck have to be considered to characterize the income determination.

2.2 Effort, motivation and the income determination: basic assumptions

The economy is populated by a continuum of individuals whose actions take place according to the timeline on Figure 2. Each individual lives for two periods: childhood and adulthood. When adults, they work in order to maximize their welfare and the consumption of their household. They also vote over income redistribution. When children, they are educated and socialized, and by this process they internalize the cultural practices which will influence their behavior when they become adults. Indeed, as noted

[5] According to Andreoni and Bernheim (2009), the 50/50 norm may also reflect that people like to be perceived as fair.

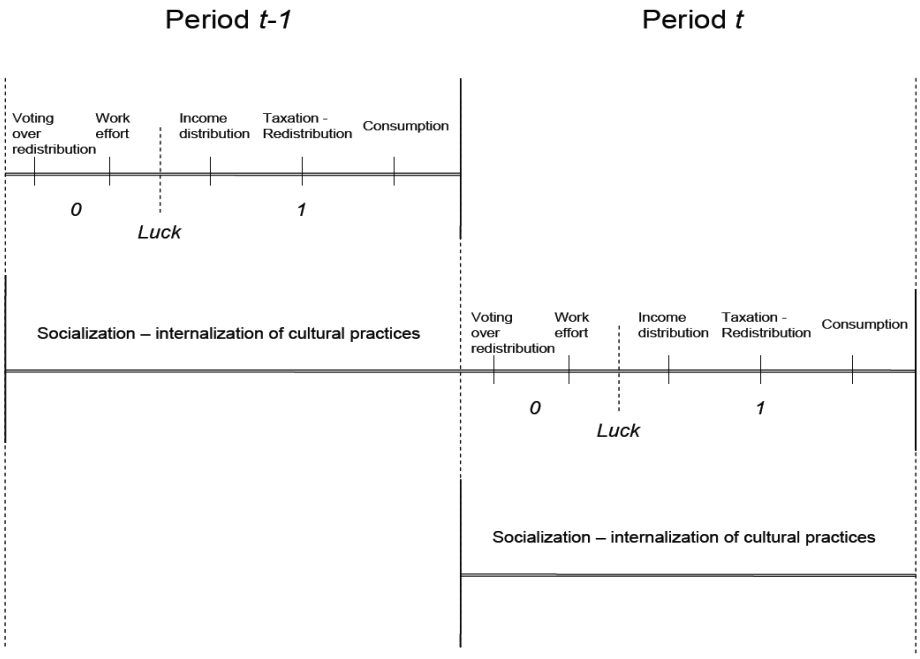

Figure 2: Timing of actions

in Alesina and Giuliano (2010), social psychologists argue that the cultural environment during youth can leave a permanent mark on individuals, while after reaching adulthood they are resistant to change. To assess the cultural side of human behavior, recent studies have pointed out the significant and persistant difference between immigrant and native behaviors such as on fertility choices and women's labor supply (Fernández and Fogli, 2006), on savings (Carroll et al., 1994), on trust (Algan and Cahuc, 2010) or on preferences for redistribution (Luttmer and Singhal, 2011, Alesina and Giuliano, 2010).

Following Piketty (1995), Alesina and Angeletos (2005) and Bénabou and Tirole (2006), we assume that income of an adult at date t is determined conjointly by luck and by effort such as:

$$y_{it} = e_{it} + \varepsilon_i \tag{1}$$

where e_{it} is her effort and ε_i represents her luck (or bad luck), unknown before the income distribution and such as $E_0[\varepsilon_i] = 0$ (see Fig. 2). When considering effort, as pointed out by Bénabou and Tirole (2002), a basic assumption in economics is that individuals respond only to external rewards. As a consequence, if income is not related to effort, no effort is made. By contrast, psychologists (see Ryan and Deci, 2000, Bénabou and Tirole, 2002) stress the importance of intrinsic motivation in making an effort, where intrinsic motivation represents incentives unrelated (or eventually negatively related) to external rewards such as income or status. Assume then that effort can be specified as:

$$e_{it} = e_{it}^{im} + e_{it}^{em} \tag{2}$$

where e_{it}^{im} and e_{it}^{em} are respectively the intrinsically and extrinsically motivated efforts. Intrinsic motivation exists directly between a person and an activity. On one hand, an effort can be said intrinsically motivated if related to an interesting activity. On the other hand, an effort can be said intrinsically motivated if it satisfies any innate psychological need or taste. In both cases, the reward is the effort itself. In this view, let us consider a taste for effort a_i distributed independently from luck such as:

$$e_{it}^{im} = a_i \tag{3}$$

In addition, assume that the extrinsically motivated effort entails a utility loss reduced by the taste for effort and equal to $\frac{\left(e_{it}^{em}\right)^2}{2a_i}$. Considering a risk neutral individual, her private utility can be then specified as:

$$u_{it} = c_{it} - \frac{\left(e_{it}^{em}\right)^2}{2a_i} \tag{4}$$

where c_{it} denotes her consumption. At each period t, income redistribution is characterized by a taxation rate τ_t and a flat-rate benefit g_t. Assuming

a balanced budget, it follows that $g_t = \tau_t \bar{y}_t$, where $\bar{y}_t$ is the mean income in t. As individuals consume all their disposable income, it yields $c_{it} = y_{it}(1-\tau_t) + \tau_t \bar{y}_t$. In addition, the optimal extrinsically motivated effort which corresponds to the maximization of the expected utility $E_0[u_{it}]$ is:

$$e_{it}^{em} = a_i (1 - \tau_t) \tag{5}$$

As redistribution lowers the market return to effort, it reduces the extrinsically motivated effort. In addition, as the taste for effort lowers the utility cost of effort, it enhances the extrinsically motivated effort. Considering (2), (3) and (5), the pre-tax income (1) of an adult in t can be rewritten as:

$$y_{it} = a_i (2 - \tau_t) + \varepsilon_i \tag{6}$$

As the level of effort is reduced by redistribution, obviously the pre-tax income is also reduced. As a consequence, redistribution reduces not only the variance of the disposable income, but also the variance of the pre-tax income. Considering here both intrinsic and extrinsic motivation is an important feature of the model. Indeed, if $\tau_t = 1$ there is no extrinsic motivation then no extrinsically motivated effort. In this case, if there was no intrinsic motivation, income would be determined only by luck. By contrast, with intrinsic motivation, individuals still continue to make different levels of effort.

2.3 The fair level of redistribution

Connected to the dual-process theory of decision making (Kahneman, 2003) and the "new synthesis" in moral psychology (Haidt, 2001, 2007, 2008; see also Nado et al., 2006), let us assume a cognitive process which automatically and quickly generates moral intuitions. In addition, assume that these intuitions regarding income distribution and effort are characterized by the

deontological principle *each person should receive what he deserves*[6], we define the fair private utility $\hat{u}_{it}$ as :

$$\hat{u}_{it} = \hat{c}_{it} - \frac{(e_{it}^{em})^2}{2a_i} \tag{7}$$

where $\hat{c}_{it}$ is the consumption allowed by the deserved or fair income $\hat{y}_{it} = e_{it}$, i.e. the income only related to effort. Following Alesina and Angeletos (2005), we then specify the universal engine of fairness by:

$$F_t = \int_i (u_{it} - \hat{u}_{it})^2 \, di \tag{8}$$

This fair-motivated cognitive process urges people to reduce unfairness. As a and ε are independently distributed, eq. (6) allows to rewrite eq. (8) as (see Appendix A):

$$\frac{F_t}{\sigma_a^2} = (1 - \tau_t)^2 L + \tau_t^2 (2 - \tau_t)^2 \tag{9}$$

where $L = \frac{\sigma_\varepsilon^2}{\sigma_a^2}$, σ_ε^2 is the variance of ε which represents the importance of luck in the income determination (one can think of for example a uniform distribution where $\varepsilon_{\max} = \pm l$, and then where $\sigma_\varepsilon^2 = \frac{l^2}{3}$) and σ_a^2 is the variance of a. L represents therefore the relative importance of luck in the income determination and the *intuitive* fair tax rate which corresponds to the minimization of (9) is as follows (see Appendix A):

$$\tau_t^f = \tau^f = \begin{cases} 1 - \sqrt{1 - \frac{L}{2}} & \text{if } L \leq 2 \\ 1 & \text{otherwise} \end{cases} \tag{10}$$

[6] Forsé et Parodi (2006) show that European countries share an identical hierarchy of moral principles: first the guarantee of basic needs, second fairness (merit), and far less important equality of income. If we admit that basic needs are mostly satisfied in Europe and in the United States, fairness is the relevant concept to analyse marginal variations of the redistribution levels. In addition, Schokkaert (1998) argues that fairness is the dominant criteria when considering social relationship at an aggregate level.

Under the fair motive, the tax rate is increased with the relative importance of luck in the income determination: $\frac{\partial \tau^f}{\partial L} \geq 0$. As the relative importance of luck is defined by $L = \frac{\sigma_\varepsilon^2}{\sigma_a^2}$, it also means that an increase of the variance of a reduces the relative luck and then $\frac{\partial \tau^f}{\partial \sigma_a^2} \leq 0$. If income was determined only by luck, eq. (10) would imply that $\tau^f = 1$, i.e. that a fair income distribution would be characterized by an equality of income as in the Ultimatum Game. One can also note that with no intrinsic motivation the only taxation consistent with the minimization of (8) is $\tau^f = 1$. Indeed, in such a case $\tau^f = 1$ means no effort and then an income distribution only characterized by luck. In $\tau^f = 1$ we would have then $F_t = 0$ which is the obvious minimum.

In addition, let us admit that an increase of the variance of a is correlated with an increase between mean and median income as in standard statistical distributions. For example, assuming a distributed according to a Log-normal of parameters $\mu = 0$ and σ yields $\sigma_a^2 = \left(e^{\sigma^2} - 1\right) e^{\sigma^2}$ and $\bar{a} - a_{med} = \Delta = e^{\frac{\sigma^2}{2}} - 1$. In this case $L = \frac{\sigma_\varepsilon^2}{\Delta(2+\Delta)(1+\Delta)^2}$ and any increase of the difference between mean and median income is associated with a lower relative importance of luck in the income determination and then with a lower intuitively fair level of redistribution. However, as highlighted in the Ultimatum Game, the fair intuition can not explain by itself the whole process of decision-making when considering the demand for redistribution.

3 Guilt aversion, rational control and the history dependancy of redistribution

As the output of an automatic and quick process, τ^f appears as an obvious (intuitive) moral imperative. We then assume that deviation from this moral reference point can generate guilt and that this feeling is stronger as the gap increases. If an individual is only driven by guilt aversion regarding

redistribution, he will always choose the fair level τ^f. However, as explained by Gray (2004), "At some point of processing, functional specialization is lost, and emotion and cognition conjointly and equally contribute to the control of thought and behavior", i.e. emotion is integrated into the rational reasoning or is controlled by this rational process.

3.1 Guilt as self-evaluative emotion

In the cognitive trend of psychology, in line with evolutionary theory, emotion is often viewed as a universal set of largely prewired internal processes of self-maintenance and self-regulation (Ekman, 1992, LeDoux, 1996). Under this biological aspect of emotion, nonverbal measures such as facial expressions, skin conductance, or the activation of certain parts of the brain (amygdala, insula, ...) have been favored to assess the universality of the emotional phenomena (see Phelps, 2009). Using facial expressions, Ekman and Friesen (1971) have for example suggested that there are six basic emotional expressions (happy, sad, fear, anger, disgust and surprise) whose characteristics are universally observed across culture.

By contrast, in the cultural trend of psychology closer to anthropology, along with these six basic emotions there are also emotions such as guilt or pride which are complex social emotions whose experience is not characterized by universality[7]. To illustrate their point, Tracy and Robins (2004) explain that a person may feel great happiness after winning either a lottery or an athletic event, but that only the athletic success can generate pride. Following their reasoning, we can add that if an athletic success can generate pride, it can also generate guilt if there is cheating or drug use, but obviously the experience of guilt will be weaker as drug use is generalized and

[7] See Markus and Kitayama (1991), Mesquita and Frijda (1992), Frijda and Mesquita (1994), Scherer (1997), Eisenberg (2000), Tracy and Robins (2004), Goetz and Keltner (2007), Edelstein and Shaver (2007).

established as a norm. In other words, guilt is context-dependent because it is based on a comparison process with others' behaviors and with cultural practices. As noted by Elster (1998), guilt is a self-evaluative emotion. Supporting such a view, Harlé and Sanfey (2007) and Twenge et al. (2007) have shown that manipulating the environment in prosocial behavior experiments (Ultimatum Game in Harlé and Sanfey, 2007) can change significantly their output. In addition, prosocial behavior experiments (among which UG) undertaken in small-scale societies suggest that "culturally transmitted behavioral variation may substantially affect decision-making" (Henrich, 2000; see also Henrich et al., 2001, 2005).

To take into account this cultural side of human behavior, let us then assume that, even if the moral intuition is universal, the self-evaluative expression of guilt is shaped by internalization of cultural practices during childhood through the process of socialization[8]. More precisely, let us assume that unfairness of the institution observed by a person during his childhood reduces his feeling of guilt when he becomes adult. We can in this case represent the self-evaluative experience of guilt G_t of adults in t by:

$$G_t = \frac{\left(\tau_t^f - \tau_t\right)^2}{\Omega_{t-1}} \tag{11}$$

where $\Omega_{t-1} = \left[\tau_{t-1}^f - \tau_{t-1}^*\right]^2$ represents the unfairness of the institution chosen by the previous adult generation in $t-1$.

3.2 Redistribution, self-control and cultural practices

Assume that guilt is integrated into the rational reasoning such as individual demands for redistribution can be characterized by the maximization of the following social utility:

[8] As stated by Bisin and Verdier (2001), internalization of cultural practices can occur through socialization inside the family and in the society via imitation and learning.

$$U_{it} = u_{it} - \frac{\Phi}{2} G_t \tag{12}$$

where Φ is a positive parameter which represents the individual weight of guilt aversion when choosing the level of redistribution. As explained in the introduction, such a specification is in the spirit of modern cognitive neurosciences (Gray, 2004, Bechara, 2004, Cohen, 2005, Pessoa, 2008). It is also consistent with Charness and Dufwenberg (2006) who have experimentally shown in the trust game that guilt aversion was shaping significantly individual behaviors. Considering (6), (4) and (11), maximizing (12) results in the following individual demands for redistribution in t:

$$\tau_{it} = \begin{cases} \frac{2(\bar{a}-a_i)+\frac{\Phi}{\Omega_{t-1}}\tau^f}{2\bar{a}-a_i+\frac{\Phi}{\Omega_{t-1}}} & \text{if } a_i \leq \bar{a} + \frac{\Phi\tau^f}{2\Omega_{t-1}} \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

Individual demands for redistribution as specified in (13) decrease with personal income and increase with the degree of unfairness of the income distribution. Eq. (13) is then consistent with empirical surveys (Fong, 2001, Corneo and Grüner, 2002, Alesina and La Ferrara, 2005, Corneo and Fong, 2008, Alesina and Giuliano, 2010). In addition, eq. (13) is consistent with findings from Luttmer and Singhal (2011) and Alesina and Giuliano (2010) which stress that culture and socialization when young determine the demand for redistribution. They indeed show that immigrants from high-preference redistribution countries continue to support higher redistribution in their destination country. In eq. (13), $\frac{\Phi}{\Omega_{t-1}}$ can be interpreted as a culturally inherited degree of collective guilt aversion. If it becomes large, individual demands for redistribution get close to the fair level.

Under the sufficient condition $\max_i \{a_i\} \leq 2\bar{a}$, preferences are single-peaked in τ and thus the median-voter theorem applies. Knowing in addition that τ_{it} as defined by eq. (13) is a decreasing function of a_i, assuming that the distribution of a is skewed to the right, i.e. $a_{med} \leq \bar{a}$, yields that the

tax rate chosen under the majority rule is specified as:

$$\tau_t^* = \frac{2\left(\bar{a} - a_{med}\right) + \frac{\Phi}{\Omega_{t-1}}\tau^f}{2\bar{a} - a_{med} + \frac{\Phi}{\Omega_{t-1}}} \tag{14}$$

Let $\tau^s = \frac{2(\bar{a}-a_{med})}{2\bar{a}-a_{med}}$ be the taxation rate chosen under the majority rule if individuals were driven only by their self interest, i.e. if $\Phi = 0$, we can rewrite (14) as:

$$\tau_t^* = \xi_t \tau^s + \left(1 - \xi_t\right)\tau^f \tag{15}$$

where

$$\xi_t = \frac{2\bar{a} - a_{med}}{2\bar{a} - a_{med} + \frac{\Phi}{\left[\tau_{t-1}^f - \tau_{t-1}^*\right]^2}} \tag{16}$$

Any taxation rate chosen under the majority rule is then expressed as a convex combination of the purely interested and the purely fair taxation rates.

Proposition 1 *Consider the tax sequence* $\{\tau_t\}_{t=0}^{t=\infty}$ *and assume that* $\tau^f - \tau^s \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$, *if* $\tau_0 \in \left]\tau^f - \delta; \tau^f + \delta\right[$ *then* $\lim_{t\to\infty} \tau_t = \tau^f$, *otherwise* $\lim_{t\to\infty} \tau_t = \frac{1}{2}\left(\tau^f + \tau^s - \sqrt{\left(\tau^f - \tau^s\right)^2 - \frac{2\Phi\tau^s}{\Delta}}\right)$ *where* $\delta = \frac{\tau^f - \tau^s - \sqrt{\left(\tau^f - \tau^s\right)^2 - \frac{2\Phi\tau^s}{\Delta}}}{2}$, $\tau^f = 1 - \sqrt{\inf\left(1 - \frac{L}{2}, 1\right)}$, $L = \frac{\sigma_\varepsilon^2}{\sigma_a^2}$, $\tau^s = \frac{2(\bar{a}-a_{med})}{2\bar{a}-a_{med}}$ *and* $\Delta = \bar{a} - a_{med}$.

The dynamics of redistribution is then history dependent. If the initial level of taxation corresponds to an institution lower but sufficiently close to the fair level, the collective degree of guilt aversion transmitted to the young generation increases. When they become adult the next period, the latter will then support a redistribution level closer to the fair level. This emotional contagion process ends with the implementation of the fair institution when the cultural degree of collective guilt aversion becomes significantly large. By contrast, starting sufficiently far from the fair level leads to a too strong

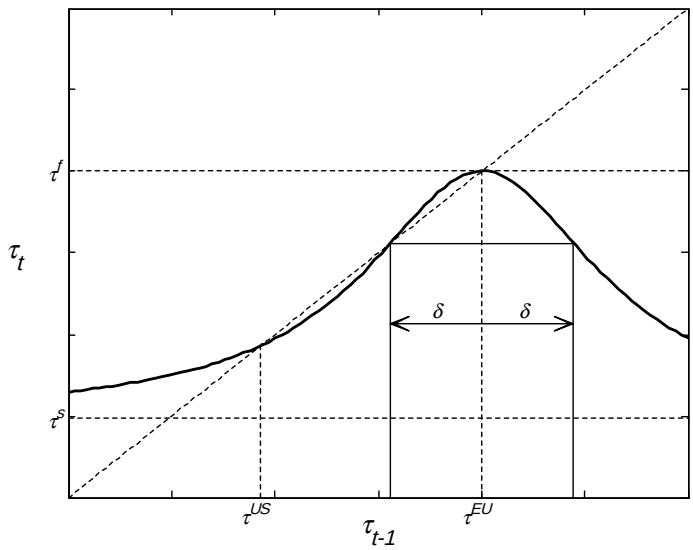

Figure 3: Multiplicity and history dependancy of redistribution when $\tau^f - \tau^s > \sqrt{\frac{2\Phi\tau^s}{\Delta}}$

signal *you should behave according to your own interest* which reduces concerns about morality and then the associated collective guilt aversion. It prevents then an emotional contagion to the benefit of a rational selfish control. The process will end with a redistribution level lower than the fair level but higher than the selfish one.

The multiplicity of steady states results from the interaction between the two distinct processes. First, guilt that people can feel related to their moral duty shapes the voting behavior: the stronger this feeling, the closer to the fair intuition their choice. Second, if people are socialized in an environment whose practices and institutions do not reflect the intuitive fairness, internalization of the observed norm *you should behave according to your own interest* will reduce individual responsability. It will then reduce the self-evaluative feeling of guilt regarding moral duty. By contrast, if people are socialized in an environment whose practices and institutions reflect

intuitive fairness, the observed norm *unfairness is intolerable* reinforce the individual feeling of guilt when choices deviate from the moral reference point. If $\tau^f - \tau^s \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$, the convergence towards the two different steady states depends therefore only on the initial level of taxation as stated in proposition 1.

According to our analysis, the high redistribution European style welfare state is then characterized by an emotional contagion whereas the low redistribution American style welfare system is characterized by a rational control: $\tau^s < \tau^{US} < \tau^{EU} = \tau^f$. As a consequence, in the European style system, as $\tau^{EU} = \tau^f$, any increase of the difference between mean and median incomes Δ can be associated with a decrease of the relative importance of luck in the wage determination L which tends to reduce the level of redistribution: $\frac{d\tau^{EU}}{d\Delta} < 0$. By contrast, in the American style system, as an increase of Δ is also associated with a higher selfish level of redistribution, $\frac{d\tau^s}{d\Delta} > 0$, the outcome is not straightforward.

Proposition 2 *If* $\tau^f - \tau^s \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$ *and* $L < 2$, $\frac{d\tau^{EU}}{d\Delta} < 0$ *while* $\frac{d\tau^{US}}{d\Delta} > 0$ *under the sufficient condition* $\varepsilon^{\Delta}_{\tau^s} \geq \frac{\tau^f - \tau^s}{\tau^f + \tau^s}$.

Under an unrestrictive condition, the American style system exhibits then a size increased with the difference between mean and median income as in the Meltzer-Richard model (1981). Such a result suggests that Europe and the United States should be dissociated when testing the empirical relationship between redistribution and income inequality. For example, while most cross-country studies (e.g. Perotti, 1996, Moene and Wallerstein, 2001, de Mello and Tiongson, 2006, Iversen and Soskice, 2006) found a negative or insignificant relation between redistribution and income inequality, Meltzer and Richard (1983) found a positive relation when considering only the United States from a time-series perspective.

4 Family background and the inheritance of inequality

So far, we have pointed out, following empirical studies, that if luck is important in the income determination, individual demands for redistribution are stronger. More generally, what is stressed in the surveys to characterize the perceived unfairness of the income distribution and then the individual demands for redistribution are factors beyond one's control (e.g. Fong, 2001, Alesina and La Ferrara, 2005, Fong et al., 2006, Corneo and Fong, 2008). Along with luck, there is then also family background. Thereafter, one can argue that luck as modelized in eq. (1) is an oversimplification of all the factors beyond one's control. Indeed, as pointed out by Bowles and Gintis (2002) and d'Addio (2007), earnings are very significantly tied to the parents' earnings[9]. When considering the income determination characterized by eq. (1), such a correlation can be obtained by assuming that psychological and cognitive skills are genetic inherited traits. However, by showing that the inherited genetic component of IQ is very weak, Bowles and Gintis (2002; see also Bisin and Topa, 2003) weaken the basis of such an assumption. By contrast, it strengthens the idea that family background is an important factor in the income determination which can not be captured by idiosyncratic shocks. In this line, Bourdieu (1984) asserts in particular that cultural consumption is of crucial importance when explaining the capacity of high-income earners to ensure the reproduction of income inequality. On

[9] As noted in Bowles and Gintis (2002), a widely held consensus among American economists used to be that "Low earnings as well as high earnings are not strongly transmitted from father to sons" (Becker, 1988). However, following Solon (1992) the low correlations between fathers' and sons' incomes previously estimated in America were due to measurement errors. With improved methodology and data, Björklund and Jäntti (1997) even show that the intergenerational income mobility in Sweden is higher (however not significantly) than in the United States.

one hand, by consuming cultural goods, high-income earners shape wealthier class cultural norms which exclude others. On the other hand, by their cultural practices, they also develop networking activities. Income, cultural capital and social capital are closely related.

On a family cultural basis of factors beyond one's control, in contrast with eq. (1), let us consider that income of an adult in t is characterized by:

$$y_{ij,t} = c_{j,t-1} + e_{it} \tag{17}$$

where $c_{j,t-1}$ represents her parents' cultural consumption in $t-1$, and $e_{i,t}$ her effort. Considering the different types of consumption goods, we redefine the private utility function as:

$$u_{ij,t} = \frac{1}{\psi^{\psi}(1-\psi)^{1-\psi}} c_{i,t}^{\psi} d_{i,t}^{1-\psi} - \frac{\left(e_{i,t}^{em}\right)^2}{2a_i} \tag{18}$$

where $d_{i,t}$ is the standard final consumption and $\psi \in [0,1]$ represents the importance of cultural goods in the consumption. Assuming for simplicity a transformation technology from one standard good to one cultural good, it yields that both forms of consumption which maximize (18) are characterized by:

$$c_{i,t} = \psi\left[(1-\tau_t)\, y_{ij,t} + \tau_t \bar{y}_t\right] \tag{19}$$

$$d_{i,t} = (1-\psi)\left[(1-\tau_t)\, y_{ij,t} + \tau_t \bar{y}_t\right] \tag{20}$$

When introducing the optimal behaviors (19) and (20) into the private utility function (18), the latter becomes: $u_{ij,t} = y_{ij,t}(1-\tau_t) + \tau_t \bar{y}_t - \frac{\left(e_{i,t}^{em}\right)^2}{2a_i}$. In this configuration, the utility maximization considering (17) leads to an unchanged level of extrinsically motivated effort compared to previous sections, i.e. $e_{i,t} = a_i(2-\tau_t)$.

Consider now that the concept of fair income is associated with the

principle of *equal opportunities* between individuals. In this case we can redefine the fair level of income as:

$$\hat{y}_{ij,t} = \bar{c}_{t-1} + e_{i,t} \tag{21}$$

where $\bar{c}_{t-1}$ is the average cultural consumption.

In line with Bowles and Gintis (2002), let us assume that personal cognitive skills characterized by a_i are not heritable genetic traits. It yields in particular that the parental cultural consumption $c_{j,t-1}$ of an individual ij is not correlated with her cognitive skills a_i. The moral intuition resulting from the minimization of (8) stays then similar to (10), i.e. $\tau_t^f = 1 - \sqrt{1 - \frac{L_t}{2}}$ if $L_t \leq 2$ ($= 1$ otherwise), except that:

$$L_t = \frac{\sigma^2_{c_{t-1}}}{\sigma^2_a} \tag{22}$$

where $\sigma^2_{c_{t-1}}$ is the variance of parents' cultural consumption in $t-1$.

Individual demands for redistribution, still characterized by the maximization of the rational utility minus guilt aversion $y_{ij,t}(1-\tau_t) + \tau_t \bar{y}_t - \frac{\left(e^{em}_{i,t}\right)^2}{2a_i} - \frac{\Phi}{2\left[\tau^f_{t-1} - \tau^*_{t-1}\right]^2}\left(\tau^f_t - \tau_t\right)^2$, are then as follows:

$$\tau_{ij,t} = \xi_t \tau^s_{ij,t} + (1-\xi_t)\,\tau^f_t \tag{23}$$

where $\tau^s_{ij,t} = \frac{2(\bar{a}-a_i)+(\bar{\pi}_{t-1}-\pi_{j,t-1})}{2\bar{a}-a_{med}}$, and $\xi_t = \frac{2\bar{a}-a_{med}}{2\bar{a}-a_{med}+\frac{\Phi}{\left[\tau^f_{t-1}-\tau^*_{t-1}\right]^2}}$. Compared with eq. (13), eq. (23) still stresses that individual demands for redistribution depend on the income inequality, the perceived unfairness of the income distribution and the specific history of the society. In addition, it stresses the importance of the family background: everything else being equal, an individual raised in a family with high cultural standards tends to support less redistribution. Following eq. (23), the tax rate chosen under the majority rule is then:

$$\tau_t^* = \xi_t \tau_t^s + (1 - \xi_t)\, \tau_t^f \tag{24}$$

where $\tau_t^s = \frac{2(\bar{a}-a_{med})+\left(\bar{\pi}_{t-1}-\pi_{med,t-1}\right)}{2\bar{a}-a_{med}}$.

Consider an institutional stationary history such as $\tau_k = \tau_{-1}, \forall k \leq t-1$. In this case we have (see Appendix D):

$$\tau_t^s = \Gamma^s(\tau_{-1}) \tag{25}$$

$$\tau_t^f = \Gamma^f(\tau_{-1}) \tag{26}$$

where $\Gamma^s(\tau_{-1}) = \frac{\left[2+\frac{\psi(1-\tau_{-1})(2-\tau_{-1})}{1-\psi(1-\tau_{-1})}\right](\bar{a}-a_{med})}{2\bar{a}-a_{med}}$ and $\Gamma^f(\tau_{-1}) = 1-\sqrt{1-\frac{\psi^2(1-\tau_{-1})^2(2-\tau_{-1})^2}{2\left[1-\psi^2(1-\tau_{-1})^2\right]}}$. From (25) and (26) we can define a purely fair and a purely self-interested tax rates corresponding respectively to $\tau^f = \Gamma^f\left(\tau^f\right)$ and $\tau^s = \Gamma^s(\tau^s)$. It results two continuous and monotonic functions $\tau^f(\psi)$ and $\tau^s(\psi, \Delta)$ such as $\frac{d\tau^f}{d\psi} \geq 0$, where $\tau^f(0) = 0$ and $\tau^f(1) = \sqrt{2}-1 \approx 0.41$, $\frac{\partial \tau^s}{\partial \psi} \geq 0$ and $\frac{\partial \tau^s}{\partial \Delta} \geq 0$, where $\tau^s(0, \Delta) = \Gamma^s(1) = \frac{2(\bar{a}-a_{med})}{2\bar{a}-a_{med}}$ and $\tau^s(\psi, 0) = 0$.

Proposition 3 *Considering a stationnary tax history* $\{\tau_k = \tau_{-1}\}_{k=-\infty}^{k=t-1}$, $\forall \psi \leq \tilde{\psi}\,(\approx 0.55)$, $\tau^f(\psi) - \tau^s(\psi, \Delta) \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$ *is a sufficient condition such as the model exhibits two local steady states characterized by* $\tau^s(\psi, \Delta) < \tau^{US} < \tau^{EU} = \tau^f(\psi)$, *where* $\tau^f(\psi) \equiv \tau^f = \Gamma^f\left(\tau^f\right)$ *and* $\tau^s(\psi, \Delta) \equiv \tau^s = \Gamma^s(\tau^s)$.

We can again assert that the possibility of a multiplicity of steady states results from the interaction of the two distinct cognitive processes. Considering family background still allows us to interpret the difference of redistribution between Europe and the United States as a distinction in the nature of the leading cognitive process when voting: affective (*system 1*) in Europe, rational (*system 2*) in the United States.

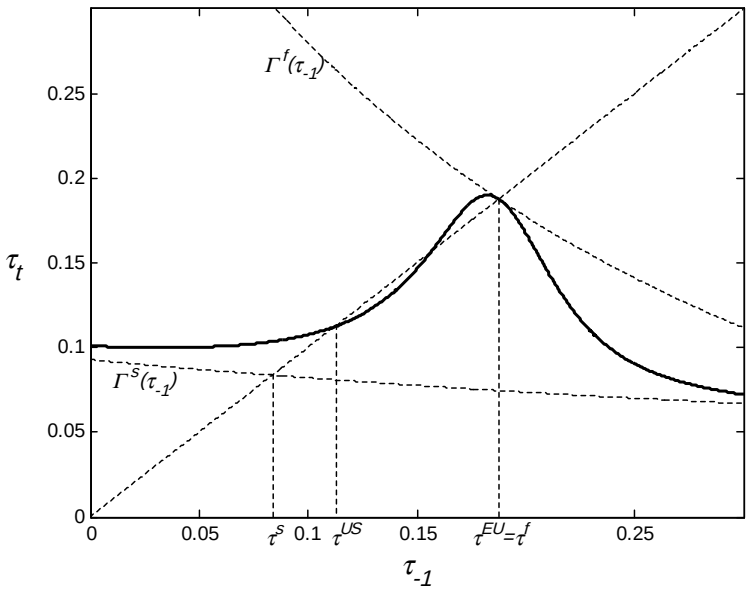

Figure 4: Multiple steady states and family background ($ln\,(a) \sim \mathcal{N}\,(0, 0.217)$, $\psi = 0.5$, $\Phi = 0.005$)

5 Conclusion

If considering that humans are only driven by their self interest, Meltzer and Richard (1981) show that the level of redistribution in a democratic society is increased with the inequality of the income distribution. A result which is weakly supported by the data. In this article, we argue that this failure of the canonical model can be associated in particular with its behavioral assumption. Modern cognitive sciences along with empirical studies converge in one major point: morality and altruism are essentials to explain redistribution. In line with the "new synthesis" in moral psychology (Haidt, 2007), we then modelize the voting behavior over redistribution as the interaction between two different mental processes. First, an automatic cognitive process which generates quick intuitions on the fair level of redistribution. Second, a rational self-oriented reasoning which controls the feeling of guilt associated with fair intuitions. As a consequence of this dual-process decision-making, and assuming that guilt aversion is context dependent and is reduced if the previous generation failed in implementing the intuitively fair institution, the model exhibits a multiplicity of steady states which can explain the huge difference of redistribution observed between Europe and the United States.

The approach we use in this article raises several issues which can lead to further research. First, the mind architecture that we use in this article, following Kahneman (2003) and Haidt (2001), states a clear partition between an emotional process and a rational one. For most neurobiologists, such a partition is a highly stylised distinction with no strong basis (see Franck et al., 2009). By contrast, in the neural workspace model proposed for example by Dehaene et al. (1998) and Dehaene and Naccache (2001), lots of distinct groups of neurons convey in parrallel different representation of the external world, and the resulting conscious perception of the external world often adopts the information of one neural group and entirely suppresses the information carried by the others according to the *winner-take-all* principle

(Camerer et al., 2005). Second, our approach suggests that emotions are in essence moral. However, as well as guilt we can think of envy and greed (see Elster, 1998) as important emotions shaping human behavior, especially when considering voting over redistribution. Finally, we have not considered individual heterogeneity in the guilt aversion. Nevertheless, following Alesina and Giuliano (2010), it appears that more educated individuals are less supportive of redistribution. This could suggest that guilt aversion is negatively correlated with cognitive skills or that more able individuals are also those who can control efficiently their emotions.

References

[1] Alesina A., Glaeser E. and Sacerdote B. (2001), Why doesn't the US have a European-style welfare system?, *Brookings Papers on Economic Activity*, 2, pp. 187-277.

[2] Alesina A. and Glaeser E. (2004), *Fighting poverty in the US and Europe: A world of difference*, Oxford University Press.

[3] Alesina A. and Angeletos G.-M. (2005), Fairness and redistribution: US versus Europe, *American Economic Review,* 95(4), pp. 960-980.

[4] Alesina A. and La Ferrara E. (2005), Preferences for redistribution in the land of opportunities, *Journal of Public Economics*, 89(5-6), pp. 897-931.

[5] Alesina A. and Giuliano P. (2010), Preferences for redistribution, forthcoming in A. Bisin and J. Benhabib eds. *Handbook of Social Economics*, North Holland Amsterdam.

[6] Algan Y. and Cahuc P. (2010), Inherited trust and growth, *American Economic Review*, 100(5), pp. 2060-2092.

[7] Anderson S., Bechara A., Damasio H., Tranel D. and Damasio A. (1999), Impairment of social and moral behavior related to early damage in human prefrontal cortex, *Nature Neuroscience*, 2, pp. 1032-1037.

[8] Andreoni J. and Bernheim D. (2009), Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects, *Econometrica*, 77(5), pp. 1607-1636.

[9] Batson C. D. (1991), *The altruism question: toward a social psychological answer*, Hillsdale, NJ: Lawrence Erlbaum.

[10] Bechara A. (2004), The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage, *Brain and Cognition*, 55, pp. 30-40.

[11] Becker G. (1988), Family economics and macro behavior, *American Economic Review*, 78(1), pp. 1-13.

[12] Bénabou R. (2000), Unequal societies: income distribution and the social contract, *American Economic Review*, 90(1), pp. 96-129.

[13] Bénabou R. and Ok E. (2001), Social mobility and the demand for redistribution: the POUM hypothesis, *Quarterly Journal of Economics*, 116(2), pp. 447-487.

[14] Bénabou R. and Tirole J. (2003), Intrinsic and extrinsic motivations, *Review of Economic Studies*, 70, pp. 489-520.

[15] Bénabou R. and Tirole J. (2006), Belief in a just world and redistributive politics, *Quarterly Journal of Economics*, 121(2), pp. 699-746.

[16] Berthoz S., Armony J., Blair R. and Dolan R. (2002), An fMRI study of intentional and unintentional (embarrassing) violations of social norms, *Brain*, 125, pp. 1696–708.

[17] Berthoz S., Grezes J., Armony J., Passingham R. and Dolan R. (2006), Affective response to one's own moral violations, *NeuroImage*, 31(2), pp. 945-950.

[18] Bisin A. and Verdier T. (2001), The economics of cultural transmission and the dynamics of preferences, *Journal of Economic Theory*, 97, pp. 298-319.

[19] Bisin A. and Topa G. (2003), Empirical models of cultural transmission, *Journal of the European Economic Association*, 1(2-3), pp. 363-375.

[20] Björklund A. and Jäntti M. (1997), Intergenerational income mobility in Sweden compared to the United States, *American Economic Review*, 87(5), pp. 1009-1018.

[21] Blair R. (2001), Neurocognitive models of aggression, the antisocial personality disorders, and psychopathy, *Journal of Neurology, Neurosurgery & Psychiatry*, 71(16), pp. 727-731.

[22] Bolton G. and Ockenfels A. (2000), ERC: A theory of equity, reciprocity, and competition, *American Economic Review*, 90(1), pp. 166-193.

[23] Bourdieu P. (1984), *Distinction: A social critique of the judgement of taste*, London: Routledge & Kegan Paul.

[24] Bowles S. and Gintis H. (2002), The inheritance of inequality, *Journal of Economic Perspective*, 16(3), pp. 3-30.

[25] Campante F. (2010), Redistribution in a model of voting and campaign contributions, *Journal of Public Economics*, forthcoming.

[26] Camerer C., Loewenstein G. and Prelec D. (2005), Neuroeconomics: how neuroscience can informe economics, *Journal of Economic Literature*, p. 9-64.

[27] Carroll C., Rhee B.-K. and Rhee C. (1994), Are there cultural effects on saving? Some cross-sectional evidence, *Quarterly Journal of Economics*, 109(3), pp. 685-699.

[28] Charness G. and Dufwenberg M. (2006), Promises and partnership, *Econometrica*, 74(6) , pp. 1579–1601.

[29] Cohen J. (2005), The vulcanization of the human brain: a neural perspective on interactions between cognition and emotion, *Journal of Economic Perspectives*, 19(4), pp. 3-24.

[30] Corneo G. and Grüner H.-P.(2002), Individual preferences for political redistribution, *Journal of Public Economics*, 83, pp. 83-107.

[31] Corneo G. and Fong C. (2008), What's the monetary value of distributive justice?, *Journal of Public Economics*, 92(1), pp. 289-308.

[32] Damasio A. (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, Penguin Books.

[33] Decety J. and Chaminade T. (2003), Neural correlates of feeling sympathy, *Neuropsychologia*, 41, pp. 127-138.

[34] Dehaene S., Kerszberg M. and Changeux J.P. (1998), A neuronal model of a global workspace in effortful cognitive tasks, *Proceedings of the National Academy of Sciences USA*, 95, pp. 14529-14534.

[35] Dehaene S. and Naccache L. (2001), Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework, *Cognition*, 79, pp. 1-37.

[36] de Mello L. et Tiongson E. (2006), Income inequality and redistributive government spending, *Public Finance Review*, 34(3), pp. 282-305.

[37] Edelstein R. and Shaver P. (2007), A cross-cultural examination of lexical studies of self-conscious emotions, in Tracy J., Robins R. and Tangney J. P. (Eds), *The self-conscious emotions: Theory and research*, NY: The Guilford Press, chap. 11.

[38] Eisenberg N. (2000), Emotion, regulation and moral development, *Annual Review of Psychology*, 51, pp. 665-697.

[39] Ekman P. and Friesen W. V. (1971), Constants across cultures in the face and emotion, *Journal of Personality and Social Psychology*, 17, pp. 124-129.

[40] Ekman P. (1992), An argument for basic emotions, *Cognition and Emotion*, 6, pp. 169-200.

[41] Elster J. (1998), Emotions and economic theory, *Journal of Economic Literature*, 36(1), pp. 47-74.

[42] Evans J. (2008), Dual-processing accounts of reasoning, judgment and social cognition, *Annual Review of Psychology*, 59, pp. 255-278.

[43] Fehr E. and Schmidt K. (1999), A theory of fairness, competition, and cooperation, *Quarterly Journal of Economics*, 114(3), pp. 817-868.

[44] Fehr E. and Schmidt K. (2006), The economics of fairness, reciprocity and altruism: experimental evidence and New Theories, in S.-C. Kolm and J. Mercier Ythier (Eds), *Handbook of the economics of giving, altruism and reciprocity, vol. 1*, North Holland/Elsevier, chap. 8.

[45] Fernández R. and Fogli A. (2006), Fertility: the role of culture and family experience, *Journal of the European Economic Association*, 4(2–3) pp. 552–561.

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

[47] Fong C., Bowles S. and Gintis H. (2006), Strong Reciprocity and the Welfare State, in S.-C. Kolm and J. Mercier Ythier (Eds), *Handbook of the economics of giving, altruism and reciprocity, vol. 2*, North Holland/Elsevier, chap. 23.

[48] Forsé M. and Parodi M. (2006), Justice distributive: la hiérachie des principes selon les européens, *Revue de l'OFCE*, 98, pp. 213-244.

[49] Franck M., Cohen M. and Sanfey A. (2009), Multiple systems in decision making, Current Directions in Psychological Science, 18(2), pp. 73-77.

[50] Frijda N. and Mesquita B. (1994), The social roles and functions of emotions, in Kitayama S. and Markus H. (Eds), *Emotion and Culture: Empirical studies of mutual influence*, Washington DC: American Psychological Association, chap. 3.

[51] Gilens M. (1999), *Why Americans hate welfare*, University of Chicago Press, Chicago.

[52] Goetz J. and Keltner D. (2007), Shifting meanings of self-conscious emotions across cultures: A social-functional approach, in Tracy J., Robins R. and Tangney J. P. (Eds), *The self-conscious emotions: Theory and research*, NY: The Guilford Press, chap. 9.

[53] Gray J. (2004), Integration of emotion and cognitive control, *Current Directions in Psychological Science*, 13, pp. 46-48.

[54] Greene J., Sommerville R., Nystrom L., Darley J., Cohen J. (2001), An fMRI investigation of emotional engagement in moral judgement, *Science*, 293, pp. 2105-2108.

[55] Greene J., Nystrom L., Engell A., Darley J., Cohen J. (2004), The moral bases of cognitive conflict and control in moral judgment, *Neuron*, 44, pp. 389-400.

[56] Greene, J. (2005), Cognitive neuroscience and the structure of the moral mind, in S. Laurence, P. Carruthers. and S. Stich. (Eds.), *The Innate Mind: Structure and Contents*, Oxford University Press, New York.

[57] Güth W., Schmittberger R. and Schwarze B. (1982), An experimental analysis of ultimatum bargaining, *Journal of Economic Behavior & Organization*, 3, p. 367-388.

[58] Haidt J. (2001), The emotional dog and its rational tail: a social intuitionist approach to moral judgment, *Psychological Review*, 108, pp. 814-834.

[59] Haidt J. (2007), The new synthesis in moral psychology, *Science*, 316(5827), p.998-1002.

[60] Haidt J. (2008), Morality, *Perspectives on Psychological Science*, 3, p.65-72.

[61] Harlé M. and Sanfey A. (2007), Incidental sadness biases social economic decisions in the Ultimatum Game, *Emotion*, 7(4), pp. 876-881.

[62] Hauser M. (2006), *Moral minds: How nature design our universal sense of right and wrong*, Harper Collins.

[63] Henrich J. (2000), Does culture matter in economic behavior? Ultimatum Game bargaining among the Machiguenga of the Peruvian Amazone, *American Economic Review*, 90(4), pp. 973-979.

[64] Henrich J., Boyd R., Bowles S., Camerer C., Gintis H., McElreath R. and Fehr E. (2001), In search of Homo economicus: Experiments in 15 small-scale societies, *American Economic Review*, 91(2), pp. 73-79.

[65] Henrich J., Boyd R., Bowles S., Gintis H., Fehr E., Camerer C., McElreath R., Gurven M., Hill K., Barr A., Ensminger J., Tracer D., Marlow F., Patton J., Alvard M., Gil-White F. and Henrich N. (2005), "Economic Man" in cross-cultural perspective: Ethnography and experiments from 15 small-scale societies, *Behavioral and Brain Studies*, 28, pp. 795-855.

[66] Hsu M., Anen C. and Quartz S. (2008), The right and the good: Distributive justice and neural encoding of equity and efficiency, *Science*, 320, pp. 1092- 1095.

[67] Iversen T. and Soskice D. (2006), Electoral institutions and the politics of coalitions: why some democracies redistribute more than others, *American Political Science Review*, 100(2), pp. 165-181.

[68] Kahneman D. (2003), Maps of bounded rationality: Psychology for behavioral economics, *American Economic review*, 93(5), pp. 1449-1475.

[69] Koenigs M. and Tranel D. (2007), Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the Ultimatum Game, *The Journal of Neuroscience*, 27(4), pp. 951-956.

[70] LeDoux J. (1996), *The Emotional Brain*, New York, Simon and Schuster.

[71] Lerner M. (1982), *The Belief in a Just World: a fundamental delusion*, New York, NY: Plenum Press.

[72] Lind J. T. (2005), Why is there so little redistribution?, *Nordic Journal of Political Economy*, 31, pp. 111-125.

[73] Lind J. T. (2007), Fractionalization and the size of government, *Journal of Public Economics*, 91(1-2), pp. 51-76.

[74] Luttmer E. and Singhal M. (2011), Culture, context, and the taste for redistribution, *American Economic Journal: Economic Policy*, 3(1), pp. 157-179.

[75] Markus H. and Kitayama S. (1991), Culture and the self: implications for cognition, emotion, and motivation, *Psychological Review*, 98(2), pp. 224-253

[76] Meltzer A. and Richard S. (1981), A rational theory of the size of government, *Journal of Political Economy*, 89(5), pp 914-927.

[77] Meltzer A. and Richard S. (1983), Tests of a rational theory of the size of government, *Public Choice*, 41, pp. 403–418.

[78] Mesquita B. and Frijda N. (1992), Cultural variations in emotions: A review, *Psychological Bulletin*, 112(2), pp. 179-204.

[79] Mikhail J. (2007), Universal moral grammar: theory, evidence and the future, *Trends in Cognitive Sciences*, 11(4), pp. 143-152.

[80] Moene K. O. et Wallerstein (2001), Inequality, social insurance and redistribution, *American Political Science Review*, 95(4), pp. 859-874.

[81] Moll J., Andreiuolo P.A., de Oliveira-Souza R., Eslinger P., Mourão-Miranda J. and Pessoa L. (2002), The neural correlates of moral sensitivity: a functional magnetic resonance imaging investigation of basic and moral emotions, *The Journal of Neuroscience*, 22(7), pp. 2730-2736.

[82] Moll J., Zahn R., de Oliveira-Souza R., Krueger F.and Grafman J. (2005), The neural basis of human moral cognition, *Nature Reviews Neuroscience*, 6, pp. 799-809.

[83] Nado J., Kelly D and Stich S. (2006), Moral judgement, in Symons J. and Calvo P. (Eds), *Routledge Companion to the Philosophy of Psychology*, Routledge, forthcoming.

[84] Perotti R. (1996), Growth, income distribution and democracy: what the data say, *Journal of Economic Growth*, 1(2), pp. 149–187.

[85] Pessoa L. (2008), On the relationship between emotion and cognition, *Nature Reviews Neuroscience*, 9, pp. 223-227.

[86] Petrova M. (2008), Inequality and Media Capture, *Journal of Public Economics*, 92(1-2), pp. 183-212.

[87] Phelps E. (2009), The study of emotion in neuroeconomics, in Glimcher P., Camerer C., Fehr E. and Poldrack R. (Eds), *Neuroeconomics: Decision making and the brain*, Academic Press, Elsevier, chap. 16.

[88] Piketty T. (1995), Social mobility and redistributive politics, *Quarterly Journal of Economics*, 110(3), pp. 551-584.

[89] Rodriguez F. (2004), Inequality, redistribution and rent-seeking, *Economics & Politics*, 16, pp. 287-320.

[90] Ryan R. and Deci E. (2000), Intrinsic and extrinsic motivations: classic definitions and new directions, *Contemporary Educational Psychology*, 25, pp. 54-67.

[91] Sanfey A., Rilling J., Aronson J., Nystrom L. and Cohen J. (2003), The neural basis of economic decision-making in the ultimatum game, *Science*, 300, pp. 1755-1758.

[92] Scherer K. (1997), The role of culture in emotion-antecedent appraisal, *Journal of Personality and Social Psychology*, 73(5), pp. 902-922.

[93] Schokkaert E. (1998), Mr. Fairmind is post-welfarist: opinions on distributive justice, *Center for Economic Studies DPS 98.09*, Catholic University of Leuven.

[94] Solon G. (1992), Intergenerational income mobility in the United States, *American Economic Review*, 82(3), pp. 393-408.

[95] Roemer J. (1998), Why do the poor do not expropriate the rich: an old argument in new garb, *Journal of Public Economics*, 70, pp. 399-424.

[96] Tracy J. and Robins R. (2004), Putting the self into self-conscious emotions: a theoretical model, *Psychological Inquiry*, 15(2), pp. 103-125.

[97] Twenge J., Baumeister R., DeWall N., Ciarocco N. and Bartels M. (2007), Social exclusion decreases prosocial behavior, *Journal of Personality and Social Psychology*, 92(1), pp. 56-66.

[98] van 't Wout M., Kahn R., Sanfey A. and Aleman A. (2006), Affective state and decision-making in the Ultimatum Game, *Experimental Brain Research*, 169, pp. 564-568.

[99] Young L. and Koenigs M. (2007), Investigating emotion in moral cognition: a review of evidence from functional neuroimaging and neuropsychology, *British Medical Bulletin*, 84, pp. 69-79.

Appendix A. The fair-oriented cognitive process

According to eqs. (4) and (7), the moral objective $F_t = \int_i (u_{it} - \hat{u}_{it})^2 \, di$ can be rewritten as:

$F_t = \int_i \{[(1-\tau_t) y_{it} + \tau \bar{y}_t] - \hat{y}_{it}\}^2 \, di$

According to eq. (6),

$F_t = \int_i \{[(1-\tau_t)(a_i(2-\tau_t) + \varepsilon_i) + \tau_t \bar{a}(2-\tau_t)] - a_i(2-\tau_t)\}^2 \, di$

$= \int_i \{(1-\tau_t)\varepsilon_i - \tau_t(2-\tau_t)(a_i - \bar{a})\}^2 \, di.$

As a and ε are independently distributed over the population,

$F_t = (1-\tau_t)^2 \int_i \varepsilon_i^2 di + \tau_t^2 (2-\tau_t)^2 \int_i (a_i - \bar{a})^2 \, di = (1-\tau_t)^2 \sigma_\varepsilon^2 + \tau_t^2 (2-\tau_t)^2 \sigma_a^2.$

It follows that

$$\frac{F_t}{\sigma_a^2} = (1-\tau_t)^2 L + \tau^2 (2-\tau_t)^2$$

where $L = \frac{\sigma_\varepsilon^2}{\sigma_\alpha^2}$.

It follows that $\frac{\partial \frac{F_t}{\sigma_a^2}}{\partial \tau_t} = 4(1-\tau_t)\left[-\tau_t^2 + 2\tau_t - \frac{L}{2}\right]$ and $\frac{\partial^2 \frac{F_t}{\sigma_a^2}}{\partial \tau_t^2} = -4\left[-\tau_t^2 + 2\tau_t - \frac{L}{2}\right] + 8(1-\tau_t)^2$.

If $L > 2$, $-\tau_t^2 + 2\tau_t - \frac{L}{2} < 0 \ \forall \tau_t \leq 1$. It follows that $\frac{\partial \frac{F_t}{\sigma_a^2}}{\partial \tau_t}(1) = 0$, $\frac{\partial^2 \frac{F_t}{\sigma_a^2}}{\partial \tau_t^2}(1) > 0$ and then $\underset{\tau_t \leq 1, L > 2}{\arg\min} F_t = 1$.

If $L \leq 2$, $-\tau_t^2 + 2\tau_t - \frac{L}{2} = -\left(1 - \sqrt{1 - \frac{L}{2}} - \tau_t\right)\left(1 + \sqrt{1 - \frac{L}{2}} - \tau_t\right)$. It follows that $\frac{\partial \frac{F_t}{\sigma_a^2}}{\partial \tau_t}(1) = 0$, $\frac{\partial^2 \frac{F_t}{\sigma_a^2}}{\partial \tau_t^2}(1) \leq 0$ while $\frac{\partial \frac{F_t}{\sigma_a^2}}{\partial \tau_t}\left(1 - \sqrt{1 - \frac{L}{2}}\right) = 0$, $\frac{\partial^2 \frac{F_t}{\sigma_a^2}}{\partial \tau_t^2}\left(1 - \sqrt{1 - \frac{L}{2}}\right) \geq 0$ and then $\underset{\tau_t \leq 1, L \leq 2}{\arg\min} F_t = 1 - \sqrt{1 - \frac{L}{2}}$.

Appendix B. Proof of proposition 1

Let us define $\delta_t = \tau^f - \tau_t^*$ the difference between the fair and the effective level of taxation, eq. (15) can be rewritten as:

$$\delta_t = \frac{2\bar{a} - a_{med}}{2\bar{a} - a_{med} + \frac{\Phi}{\delta_{t-1}^2}} \left(\tau^f - \tau^s\right) \tag{27}$$

and the stationnarity is then defined by:

$$\delta^3 - \left(\tau^f - \tau^s\right)\delta^2 + \frac{\Phi}{2\bar{a} - a_{med}}\delta = 0 \tag{28}$$

If $\left|\tau^f - \tau^s\right| \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$, eq. (28) exhibits three roots $\delta = 0$, $\delta = \frac{\tau^f - \tau^s + \sqrt{\left(\tau^f - \tau^s\right)^2 - \frac{2\Phi\tau^s}{\Delta}}}{2}$ and $\delta = \frac{\tau^f - \tau^s - \sqrt{\left(\tau^f - \tau^s\right)^2 - \frac{2\Phi\tau^s}{\Delta}}}{2}$.

In addition, as $\lim_{\delta^2 \to 0} \frac{\partial\left[\frac{2\bar{a}-a_{med}}{2\bar{a}-a_{med}+\frac{\Phi}{\delta^2}}\right]}{\partial\delta^2} = 0$, if $\tau^f > \tau^s$, there exists two steady states characterized by $\delta = 0$ and $\delta = \frac{\tau^f - \tau^s + \sqrt{\left(\tau^f - \tau^s\right)^2 - \frac{2\Phi\tau^s}{\Delta}}}{2}$, i.e. respectively by $\tau^* = \tau^f$ and $\tau^* = \frac{1}{2}\left(\tau^f + \tau^s - \sqrt{\left(\tau^f - \tau^s\right)^2 - \frac{2\Phi\tau^s}{\Delta}}\right)$, where $\tau^f = \begin{cases} 1 - \sqrt{1 - \frac{L}{2}} & \text{if } L \leq 2 \\ 1 & \text{otherwise} \end{cases}$, $L = \frac{\sigma_\varepsilon^2}{\sigma_a^2}$, $\tau^s = \frac{2(\bar{a}-a_{med})}{2\bar{a}-a_{med}}$ and $\Delta = \bar{a} - a_{med}$.

Appendix C. Proof of proposition 2

As far as $\tau^f - \tau^s \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$, the American style system is characterized by the following payroll tax:

$$\tau^{US} = \frac{1}{2}\left(\tau^f + \tau^s - \sqrt{\left(\tau^f - \tau^s\right)^2 - \frac{2\Phi\tau^s}{\Delta}}\right)$$

We have then:

$$\begin{aligned} \frac{\partial\tau^{US}}{\partial\Delta} &= \frac{1}{2}\left(\frac{\partial\tau^f}{\partial\Delta} + \frac{\partial\tau^s}{\partial\Delta} - \frac{2\left(\tau^f - \tau^s\right)\left(\frac{\partial\tau^f}{\partial\Delta} - \frac{\partial\tau^s}{\partial\Delta}\right) - 2\frac{\Phi}{\Delta}\frac{\partial\tau^s}{\partial\Delta} + 2\frac{\Phi}{\Delta}\frac{\tau^s}{\Delta}}{2\sqrt{\left(\tau^f - \tau^s\right)^2 - \frac{2\Phi\tau^s}{\Delta}}}\right) \\ &= \frac{1}{2}\left(1 - \frac{\tau^f - \tau^s}{\sqrt{\left(\tau^f - \tau^s\right)^2 - \frac{2\Phi\tau^s}{\Delta}}}\right)\frac{\partial\tau^f}{\partial\Delta} + \frac{1}{2}\frac{\partial\tau^s}{\partial\Delta} + \frac{\left(\tau^f - \tau^s\right)\frac{\partial\tau^s}{\partial\Delta} + \frac{\Phi}{\Delta}\frac{\partial\tau^s}{\partial\Delta} - \frac{\Phi}{\Delta}\frac{\tau^s}{\Delta}}{\sqrt{\left(\tau^f - \tau^s\right)^2 - \frac{2\Phi\tau^s}{\Delta}}} \end{aligned} \tag{29}$$

where $\frac{\partial\tau^f}{\partial\Delta} \leq 0$ and $\frac{\partial\tau^s}{\partial\Delta} \geq 0$.

From (29), we deduce that a sufficient condition such that $\frac{\partial \tau^{US}}{\partial \Delta} \geq 0$ is:

$$\left(\tau^f - \tau^s\right) \frac{\partial \tau^s}{\partial \Delta} + \frac{\Phi}{\Delta} \frac{\tau^s}{\Delta} \left(\frac{\partial \tau^s}{\partial \Delta} \frac{\Delta}{\tau^s} - 1\right) \geq 0 \tag{30}$$

Obviously, if $\varepsilon_{\tau^s}^{\Delta} = \frac{\partial \tau^s}{\partial \Delta} \frac{\Delta}{\tau^s} \geq 1$ condition (30) always holds and then $\frac{\partial \tau^{US}}{\partial \Delta} \geq 0$.

If $\varepsilon_{\Delta}^{\tau^s} < 1$, we can deduce from $\tau^f - \tau^s \geq 2\sqrt{\frac{\Phi}{2\bar{a} - a_{med}}} \Leftrightarrow \frac{\Phi}{\Delta} \leq \frac{\left(\tau^f - \tau^s\right)^2}{2\tau^s}$ that $\left(\tau^f - \tau^s\right) \frac{\partial \tau^s}{\partial \Delta} + \frac{\left(\tau^f - \tau^s\right)^2}{2\tau^s} \frac{\tau^s}{\Delta} \left(\varepsilon_{\Delta}^{\tau^s} - 1\right) \geq 0 \Rightarrow \left(\tau^f - \tau^s\right) \frac{\partial \tau^s}{\partial \Delta} + \frac{\Phi}{\Delta} \frac{\tau^s}{\Delta} \left(\varepsilon_{\Delta}^{\tau^s} - 1\right) \geq 0$. We can therefore show that condition (30) is satisfied when $\left(2 + \frac{\tau^f - \tau^s}{\tau^s}\right) \varepsilon_{\Delta}^{\tau^s} \geq \frac{\tau^f - \tau^s}{\tau^s} \Leftrightarrow \varepsilon_{\Delta}^{\tau^s} \geq \frac{\tau^f - \tau^s}{\tau^f + \tau^s}$.

Appendix D. Family background with stationary history

From (17), (20) and (2) we have:

$$c_{i,t} = \psi \left\{ (1 - \tau_t) \left[c_{j,t-1} + (2 - \tau_t) a_i \right] + \tau_t \left[\bar{c}_{t-1} + (2 - \tau_t) \bar{a} \right] \right\} \tag{31}$$

Assuming that offsprings' cognitive skills are randomly determined, independently of their parents' cognitive skills, entails also they are not correlated with their parents' cultural consumption. In such a case, when considering an institutional stationary history such as $\tau_s = \tau_{-1}, \forall s \leq t-1$, it follows from (31) that $\sigma_{c_{t-1}}^2 = \sigma_c^2 (\tau_{-1}) = \psi^2 (1 - \tau_{-1})^2 \left[\sigma_c^2 (\tau_{-1}) + (2 - \tau_{t-1})^2 \sigma_a^2\right]$, or equivalently:

$$\sigma_c^2 (\tau_{-1}) = \frac{\psi^2 (1 - \tau_{-1})^2 (2 - \tau_{-1})^2}{1 - \psi^2 (1 - \tau_{-1})^2} \sigma_a^2$$

and then following (8) and (22) that:

$$\tau_t^f = \Gamma^f (\tau_{-1}) = 1 - \sqrt{1 - \frac{\psi^2 (1 - \tau_{-1})^2 (2 - \tau_{-1})^2}{2\left[1 - \psi^2 (1 - \tau_{-1})^2\right]}}$$

It follows that:

$$\frac{\partial \Gamma^f}{\partial \tau_{-1}} = -\frac{\psi^2 (1-\tau_{-1})(2-\tau_{-1})\left[1+(1-\tau_{-1})\left(2-\psi^2(1-\tau_{-1})^2\right)\right]}{2\left(1-\tau_t^f\right)\left[1-\psi^2(1-\tau_{-1})^2\right]^2} \leq 0 \quad (32)$$

$$\frac{\partial \Gamma^f}{\partial \psi^2} = \frac{(1-\tau_{-1})^2(2-\tau_{-1})^2}{4\left(1-\tau_t^f\right)\left[1-\psi^2(1-\tau_{-1})^2\right]^2} \geq 0$$

Verifying in addition that $\lim_{\tau=0}\left[\tau-\Gamma^f(\tau)\right] = -1+\sqrt{\sup\left(1-\frac{2\psi^2}{1-\psi^2},0\right)} \leq 0$ and $\lim_{\tau=1}\left[\tau-\Gamma^f(\tau)\right] = 1$, we can derive that $\tau^f(\psi) \equiv \tau^f = \Gamma^f\left(\tau^f\right)$ is a monotonic and continuous function such as $\frac{d\tau^f}{d\psi} \geq 0$, $\tau^f(0) = 0$ and $\tau^f(1) = \sqrt{2}-1 \approx 0.41$ (see Fig. 5). We can also verify that solving $\tau^f = \Gamma^f\left(\tau^f\right)$ is equivalent to $\left(\tau^f\right)^3 - \left(3+\frac{2}{\psi^2}\right)\tau^f + 2 = 0$. Using the Cardano's formula, the general specification of $\tau^f(\psi)$ is under its trigonometric form as follows:

$$\tau^f(\psi) = 2\sqrt{1+\frac{2}{3\psi^2}}\cos\left[\frac{1}{3}\arccos\left(-\sqrt{\frac{27}{\left(3+\frac{2}{\psi^2}\right)^3}}\right)+\frac{4}{3}\pi\right] \quad (33)$$

From (31) we also have

$\bar{\pi}_{t-1} = \bar{\pi}(\tau_{-1}) = \frac{\psi}{1-\psi(1-\tau_{-1})}\left\{(1-\tau_{-1})(2-\tau_{-1})\bar{a}+\tau_{-1}\left[\bar{\pi}(\tau_{-1})+(2-\tau_{-1})\bar{a}\right]\right\}$

and

$\pi_{medt-1} = \pi_{med}(\tau_{-1}) = \frac{\psi}{1-\psi(1-\tau_{-1})}\left\{(1-\tau_{-1})(2-\tau_{-1})a_{med}+\tau_{-1}\left[\bar{\pi}(\tau_{-1})+(2-\tau_{-1})\bar{a}\right]\right\}$.

It follows that:

$$\bar{\pi}_{t-1} - \pi_{med,t-1} = \bar{\pi}(\tau_{-1}) - \pi_{med}(\tau_{-1}) = \frac{\psi(1-\tau_{-1})(2-\tau_{-1})}{1-\psi(1-\tau_{-1})}(\bar{a}-a_{med})$$

and therefore:

$$\tau_t^s = \frac{2(\bar{a}-a_{med})+(\bar{\pi}_{t-1}-\pi_{med,t-1})}{2\bar{a}-a_{med}} = \Gamma^s(\tau_{-1}) = \frac{\left[2+\frac{\psi(1-\tau_{-1})(2-\tau_{-1})}{1-\psi(1-\tau_{-1})}\right](\bar{a}-a_{med})}{2\bar{a}-a_{med}}$$

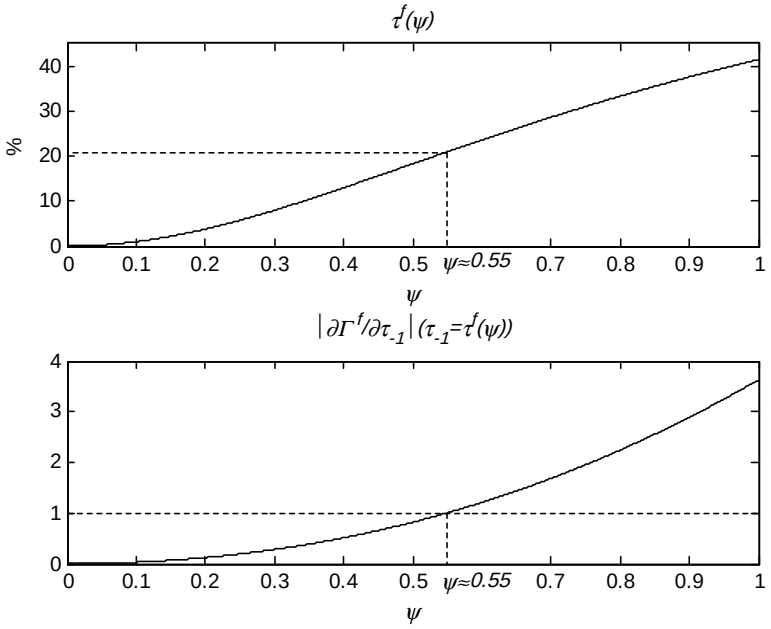

Figure 5: Characteristics of $\tau^f(\psi) \equiv \tau^f = \Gamma^f\left(\tau^f\right)$

It follows that

$$\begin{aligned}
\frac{\partial \Gamma^s}{\partial \tau_{-1}} &= -\frac{1+(1-\tau_{-1})\left[2-\psi\left(1-\tau_{-1}\right)\right]}{2\left[1-\psi\left(1-\tau_{-1}\right)\right]^2}\psi\Gamma^s(1) \leq 0 \\
\frac{\partial \Gamma^s}{\partial \psi} &= \frac{(1-\tau_{-1})(2-\tau_{-1})}{2\left[1-\psi\left(1-\tau_{-1}\right)\right]^2}\Gamma^s(1) \geq 0 \\
\frac{\partial \Gamma^s}{\partial \Delta} &= \left[1+\frac{\psi\left(1-\tau_{-1}\right)(2-\tau_{-1})}{2\left[1-\psi\left(1-\tau_{-1}\right)\right]}\right]\frac{\partial \frac{2(\bar{a}-a_{med})}{2\bar{a}-a_{med}}}{\partial \Delta} \geq 0
\end{aligned}$$

and it results a continuous and monotonic function $\tau^s(\psi,\Delta)$ such as $\frac{\partial \tau^s}{\partial \psi} \geq 0$ and $\frac{\partial \tau^s}{\partial \Delta} \geq 0$, defined by:

$$\tau^s(\psi,\Delta) = \frac{\sqrt{\left[(1-\psi)\bar{a}+(\bar{a}-a_{med})\right]^2+\psi\bar{a}^2(\bar{a}-a_{med})}-\left[(1-\psi)\bar{a}+(\bar{a}-a_{med})\right]}{2\psi\bar{a}}$$

Appendix E. Proof of proposition 3

As for proposition 1, let us define $\delta_t = \tau_t^f - \tau_t^*$ the difference between the fair and the effective level of taxation, eq (24) can be rewritten as:

$$\delta_t = \frac{2\bar{a}-a_{med}}{2\bar{a}-a_{med}+\frac{\Phi}{\delta_{t-1}^2}}\left(\tau_t^f-\tau_t^s\right) \tag{34}$$

Stationnarity is then still defined as in eq. (28), i.e. $\delta^3-\left(\tau^f-\tau^s\right)\delta^2+\frac{\Phi}{2\bar{a}-a_{med}}\delta = 0$, where $\tau^f = \Gamma^f(\tau^*) = 1-\sqrt{1-\frac{\psi^2(1-\tau^*)^2(2-\tau^*)^2}{2\left[1-\psi^2(1-\tau^*)^2\right]}}$ and $\tau^s = \Gamma^s(\tau^*) = \frac{\left[2+\frac{\psi(1-\tau^*)(2-\tau^*)}{1-\psi(1-\tau^*)}\right](\bar{a}-a_{med})}{2\bar{a}-a_{med}}$. It follows that if $\tau^f-\tau^s \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$, eq. (28) exhibits three roots of which $\delta = 0$, i.e. $\tau^* = \tau^f(\psi)$.

As $\frac{\partial \Gamma^f}{\partial \tau^*} \leq 0$ and $\frac{\partial \Gamma^s}{\partial \tau^*} \leq 0$, knowing from (24) that $\tau^* \in \left]\tau^s(\psi,\Delta),\tau^f(\psi)\right]$, it entails that $\tau^s \leq \tau^s(\psi,\Delta)$ and $\tau^f \geq \tau^f(\psi)$, and then that:

$$\tau^f(\psi)-\tau^s(\psi,\Delta) \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}} \Longrightarrow \tau^f-\tau^s \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$$

where $\tau^f(\psi) \equiv \tau^f = \Gamma^f\left(\tau^f\right)$ and $\tau^s(\psi,\Delta) \equiv \tau^s = \Gamma^s(\tau^s)$.

In addition, from (32) we can verify that $\left|\frac{\partial \Gamma^f}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^f(0)} = 0$ whereas $\left|\frac{\partial \Gamma^f}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^f(1)} = \frac{(3-\sqrt{2})\left[1+(2-\sqrt{2})\left(2-(2-\sqrt{2})^2\right)\right]}{2\left[1-(2-\sqrt{2})^2\right]^2} \approx 3.6$. It follows that $\tilde{\psi} \in]0,1[$ exists such as $\forall \psi \leq \tilde{\psi}$ $\left|\frac{\partial \Gamma^f}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^f(\psi)} \leq 1$ and then $\left|\frac{\partial \tau_t^*}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^f(\psi)=\tau^{EU}} \leq 1$. From (32) and (33) we can then compute $\tilde{\psi} \approx 0.55$ (see Fig. 5).

At last, considering $\tau^{US} \equiv \sup\left\{\delta, \delta^3 - \left(\tau^f - \tau^s\right)\delta^2 + \frac{\Phi}{2\bar{a}-a_{med}}\delta = 0\right\}$ where $\tau^f(\psi) - \tau^s(\psi,\Delta) \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$, $\left.\frac{\partial \tau_t^*}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^{US}}$ can be either positive or negative. If $\left.\frac{\partial \tau_t^*}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^{US}} \geq 0$, knowing from (24) that $\tau_t^*|_{\tau_{-1}=0} > 0$, by definition $\frac{\partial \tau_t^*}{\partial \tau_{-1}}_{\tau_{-1}=\tau^{US}} \leq 1$ if $\delta^3 - \left(\tau^f - \tau^s\right)\delta^2 + \frac{\Phi}{2\bar{a}-a_{med}}\delta = 0$ exhibits three roots.

If $\left.\frac{\partial \tau_t^*}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^{US}} \leq 0$, we can deduce from (24) and (25) that $\left|\frac{\partial \tau_t^*}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^{US}} \leq \left|\frac{\partial \Gamma^s}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^{US}}$. In addition, as $\frac{\partial^2 \Gamma^s}{\partial \tau_{-1}^2} = \frac{[1-\psi(1-\tau_{-1})]^2+\psi[1+(1-\tau_{-1})(2-\psi(1-\tau_{-1}))]}{[1-\psi(1-\tau_{-1})]^3}\psi\Gamma^s(1) \geq 0$, $\left|\frac{\partial \Gamma^s}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^{US}} \leq \left|\frac{\partial \Gamma^s}{\partial \tau_{-1}}\right|_{\tau_{-1}=0}$ where $\left|\frac{\partial \Gamma^s}{\partial \tau_{-1}}\right|_{\tau_{-1}=0} = \psi\frac{3-\psi}{2(1-\psi)^2}\Gamma^s(1)$. As $\frac{\partial \Gamma^s}{\partial \tau_{-1}} \leq 0$ (see Appendix D), $\tau^f(\psi) - \tau^s(\psi,\Delta) \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$ implies in particular that $\Gamma^s(1) \leq \tau^f(\psi)$ $\forall\psi$. It follows that $\sup\limits_{\psi\leq\tilde{\psi}}\left(\left|\frac{\partial \Gamma^s}{\partial \tau_{-1}}\right|_{\tau_{-1}=0}\right) \leq \left\{\psi\frac{3-\psi}{2(1-\psi)^2}\tau^f(\psi)\right\}_{\psi=\tilde{\psi}} \approx 0.7$, where $\tau^f\left(\tilde{\psi}\right) \approx 0.21$, and then that $\psi \leq \tilde{\psi}$ if $\left.\frac{\partial \tau_t^*}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^{US}} \leq 0$ implies $\left|\frac{\partial \tau_t^*}{\partial \tau_{-1}}\right|_{\tau_{-1}=\tau^{US}} < 1$.

It follows that, considering a stationnary tax history $\{\tau_k = \tau_{-1}\}_{k=-\infty}^{k=t-1}$, $\tau^f(\psi) - \tau^s(\psi,\Delta) \geq 2\sqrt{\frac{\Phi}{2\bar{a}-a_{med}}}$ $\forall\psi \leq \tilde{\psi} < 1$ is a sufficient condition such as the model exhibits two local steady states characterized by $\tau^s(\psi,\Delta) < \tau^{US} < \tau^{EU} = \tau^f(\psi)$.