Belief in a Just World and Redistributive Politics

Roland Bénabou
Princeton University

Jean Tirole
Institut d'Economie Industrielle

August, 2002

Preliminary and incomplete
Please do not circulate

1 Introduction

International surveys reveal striking differences between the views held in different countries concerning the causes of economic success or poverty , the extent to which individuals are responsible for their own fate, and the long-run rewards to personal effort. American "exceptionalism", as manifested by the widely professed belief in the American Dream, is but the most striking example of this phenomenon. At the same time, ethnographic surveys by sociologists reveal that working-class and lower-middle class individuals do not adhere to these views as dispassionate statisticians. On the contrary, they constantly struggle with the cognitive dissonance required to maintain (and pass on to their children) the view that effort, hard work, and good deeds will ultimately bring a better life, that crime does not pay, etc., in spite of recurrent evidence that life may not be that fair. Relatedly, experimental psychologists have documented the fact that most people have a strong need to believe that they live in a world that is just, in the sense that people generally get what they deserve, and deserve what they get. When confronted with data that contradicts this view they try hard to ignore, reinterpret, distort, or forget it –for instance by finding imaginary merits to the recipients of fortuitous rewards, or assigning blame to innocent victims.

This paper proposes a model of why people may feel such a need to believe in a just world; of why this need, and therefore the prevalence of the belief, may vary considerably across countries; and of its implications for redistributive policies (taxes and welfare payments), and the stigma born the poor. At the heart of the model are general-equilibrium interactions between each individual's psychologically-based "demand" for a belief in a just world (or similar ideology) and the degree of redistribution chosen by the polity.

Because of their imperfect willpower, individuals constantly strive to motivate themselves (or their children) towards effort, educational investment, perseverance in the face of adversity, and away from the slippery slope of idleness, welfare dependency, crime, drugs, etc. (This is another recurrent finding from the sociological evidence). In such circumstances, maintaining somewhat "rosy" beliefs about the fact that everyone will ultimately get their "just deserts" can be very valuable. Furthermore, if enough individuals end up with the view that economic success is highly dependent on effort, they will ultimately represent a pivotal voting block, and set a low tax rate. Conversely, when individuals anticipate that society will carry out little redistribution, the costs of a deficient motivation to effort or savings are much higher than with high taxes and a generous safety net. Each individual thus has greater incentives to maintain his belief that effort ultimately pays, and consequently more voters end up with such a world view.

Due to these complementarities between individual's desired beliefs, or ideological choices, and aggregate political outcomes, there can thus be two equilibria. The first one is characterized by a high prevalence of the "belief in a just world" among the population (a high degree of repression or denial of bad news about the world), and a relatively laissez-faire public policy; both are mutually sustaining. The other equilibrium is characterized by more "realistic pessimism" (less collective denial, leading to a more cynical majority), and a more generous welfare state, which in turn reduces the need to for individuals to invest in optimistic beliefs. In this equilibrium there is also less stigma born by the poor, in the sense that fewer agents are likely to blame poverty on a lack of effort or willpower.

While the "American dream" equilibrium typically involves more reality distortion –more overestimation of the extent to which people get what they deserve, can go from rags-to riches, or become president– it is also not "just" a dream, since *net* incomes or rewards are truly more closely tied to merit than in a more redistributive "realistic pessimism" equilibrium. Furthermore, this (endogenously) shared ideology has important growth and ex ante welfare benefits, since it improves individuals' deficient motivation to effort. Its net value to the poor is much more ambiguous, since they receive less transfers, and are more likely to be stigmatized.

2 Motivation / Discussion

• Why do we see such marked differences in the extent of redistribution / the "welfare state" across countries?

• American exceptionalism (noted since Toqueville; has long despaired Marxists). But also other notable cross-country differences.

• What limits the extent of redistribution in a democracy?

• Why does ideology ("right/ left wing"; emphasis on self-reliance, personal responsibility, or on "societal" causes) vary so much across countries?

I - Economists

Three main types of explanations:

1. Different beliefs: about the costs/ benefits of redistribution / the mobility process: Hirschman (1973), Piketty (1995, 1998), Bénabou and Ok (2001), Alesina and La Ferrara (2001). Or about the accuracy of employers' estimates of worker productivity: Rotemberg (2002) .

2. Multiple politico-economic equilibria: Bénabou (2000), Saint Paul (2001), Hassler-Rodriguez-Mora-Storesjletten-Zillibotti (2000, 2002), Vindigini (2002).

3. Other politico-economic explanations (e.g., exogenous institutional differences): see Alesina-Glaeser-Sacerdote (2001) for review + tests.

• Will focus here on differences in *beliefs* about the links between effort / investment and rewards. Striking differences:

– Data from World Values Survey (see Alesina et al. (2001), and Keely (2002)) shows that only 29% of Americans believe that the poor are trapped in poverty, and 30% that luck (rather than effort or education) determines income. The figures for Europeans are nearly double: 60%

and 54% respectively. Similarly, Americans are more than twice as likely as Europeans to think that the poor are lazy (60% versus 26%). Indeed, 59% of Americans agree or strongly agree that "in the long run, hard work usually brings a better life"; this view commands much less support in Europe, ranging from the low 34% in Sweden to 43% in Germany (Ladd and Bowman (2001)).

– Large differences in attitudes also exist between European countries, and particularly between the OECD and Eastern European countries: the percentage who agree or strongly agree that "in your country, people get rewarded for effort" is 36.4% in the former, and only 13.1% in the latter; the corresponding numbers for the statement that "in your country, people get rewarded for intelligence and skills" are 46% and 20% respectively (Suhrcke (2001)).

• Why such huge differences? Also, want to raise the apparently overlooked issue of whether people interpret and answer the question in a pre- or post-redistribution sense.

– Traditional Marxist explanation: workers have "false consciousness"; victims of propaganda / brainwashing by capitalists, who control education, media, etc.

– Related: coalition of the poor split on other issues (Roemer (1998)).

– Piketty : people / countries "accidentally" stuck with wrong beliefs: because costly to learn the returns to effort, at some point they stop experimenting (bandit problem).

But...

II- Sociologists – Political Scientists

Lane (1959), Hochschild (1991), Lamont (2001): very detailed interviews with hundreds of working class / moderate income workers (Black and White). Consistently find:

1a.– Obstinately / desperately cling to a belief that effort, hard work, good deeds will ultimately pay off: *people get what they deserve.* Conversely, what they get, they must deserve (good or bad).

1b – At the same time, some recognition that world is not so just; constant struggle with this "cognitive dissonance".

[Maria, cleaning lady]. *"Once Maria wonders if executives deserve their $60,000 annual salary: "I don't think they do all that [much] work, do you? Sit at their desk –they got it easy". But she suppresses the thought immediately" "Well, maybe it is a lot of work. Maybe they have a lot of writing to do, or they have to make sure things go right. So maybe they are deserving of it".* (Hochschild (1996)).

[De Angelo, machine operator] *"Personally, I think taxes are too hard. I mean a man makes, let's say* $150,000. *Well, my God, he has to give up half of that to the government –which I don't think is right. For instance if a man is fortunate enough to win the Irish Sweepstakes, he gets 150 –I think he has about* $45,000 *left. I don't think that's right".* (Lane (1959)). [See experiments on fortuitous rewards later on].

2a – Key challenge of their lives: struggle to "keep it going," not give up, persevere in the face of adversity (otherwise: welfare, homelessness, drugs...)

2b– Very harsh judgements on the (very) poor / welfare recipients (especially Blacks); poverty attributed in large part to "giving up", not caring, no "values", no direction in life. "General view that success is a triumph of the will and a reflection of ability". (Lane (1959)).

[Vincent, periodically unemployed unskilled worker]. *"If a person keeps his and works and works, and he's banking it, good luck to him! That's good. I wish to hell I could do it. I always said for year, 'I wanna get rich, I wanna get rich.' But then phew! My mind doesn't have the strong will. I say, 'Well, I'm gonna do it' Only the next day is different".* He believes that willpower is as essential as hard work to success; he has done plenty of work, but woefully lack the will. (Hochschild (1991)).

Closely echoed (we think) by psychologists...

III-Psychologists

A. "Fundamental attribution error" (Nisbett and Ross): excessive tendency to explain behavior of others by "disposition" (personal actions or attributes) rather than circumstances (luck, etc.).

B. "Illusion of control", overconfidence: excessive beliefs that they (and others), have control over their environment.

C. Belief in a Just World (Lerner (1982), Peplau and Tyler (1975)): "Individuals have a *need* to believe that they live in a world where people generally get what they deserve". Nearly universal tendency to want to believe that the world is just. When confronted with contradictory information, people try hard to ignore, reinterpret, distort, forget it. Many experiments:

1) reinterpretation of fortuitous rewards

2) Helping or blaming the victim; seeing compensating differentials. Even self-blame by victims.

3) Measurement and correlates of BJW:

– The BJW scale:

– Vietnam draft lottery experiment: those in top third of BJW scores more likely to denigrate / resent those in the group who drew an unfavorable number. Also: those who drew bad numbers lowered their own self-esteem, regardless of their BJW.

– High BJW scorers more likely to give stiff sentences to defendants convicted of a crime such as negligent homicide, but also to find the victims (e.g., in rape case) more culpable and "deserving" of their fate.

– University students solicited to volunteer up for "useful" experiments, or to 5 hours read to a blind student in high school: at normal time(volunteer up to 5 hours read tor a blind student in night school). No differences between high/low BJW scorers when appeal made 4/5 weeks before midterm fail exams. But when made 1 week / a few days before, high BJW scorers much more likely to participate / volunteer ("appeasing the gods").

– High BJW scorers: tendency to see political events in a positive sense, the status quo as desirable, politically and economically conservative, belief in active God, tendency to justify the plights of Blacks and women, negative correlation to measures of activism. Less cynical. (Peplau and Tyler (1975)):

– Correlation with Protestant ethic, belief in internal locus of control.

– French vs. American workers; the French seem to have much less BJW (Lamont (2001)).

– Where do economists fit?

Questions

– Why do people want / "need" to believe in a just world?

– To what extent can they succeed in achieving such beliefs "false consciousness" (if the word is not-so-just)?

– Why are there such variations in BJW across countries? (Also groups, individuals).

– What are the political economy implications of BJW: redistribution / welfare, stigma on the poor, etc. And others.

Proposed answer

Theory based on

1. *Imperfect willpower* $\Rightarrow$ need to motivate oneself towards effort, educational, not giving up, etc. "Rosy" beliefs about people ultimately getting their "just deserts" can be functional. (Alternative: reassuring view of the world, reduces anxiety).

2. *Motivated beliefs: rational self-deception* through endogenously selective memory / attention/ awareness.

3. *General equilibrium interactions* between individual's "demand" for BJW and the degree of redistribution chosen by the polity (i.e., between psychology/ ideology side, and political economy side):

• Many people think success is highly dependent on effort => ceteris paribus, majority or pivotal group will want a low relatively tax rate.

• People anticipate a low tax rate: their fate is highly dependent on their effort $\Rightarrow$ the costs of insufficient motivation / procrastination are higher (no safety net / welfare state). Conversely, the rewards to being highly motivated are higher $\Rightarrow$ individual have greater incentives to maintain BJW.

Thus, two equilibria:

• *Laissez-faire / BJW:* equilibrium: high degree of collective repression, "learned optimism"; little redistribution. Mutually sustaining. Also implies "blaming" poverty on lack of effort or willpower by the poor.

• *Welfare state / Realistic-Pessimist:* low degree of repression , high redistribution, generous welfare state. More "understanding" of the poor.

Is the "American dream" just a dream?

i) Yes and no: yes in the sense that more overestimation of the extent to which "people get what they deserve", can go from rags-to riches, the poor are not trapped, everyone can become president, etc., No in the sense that *net* incomes / rewards are truly more closely tied to "merit" in a BJW equilibrium. .

ii) May be a very useful illusion / ideology: higher motivation / effort, higher aggregate output / growth, etc. Less clear for the poor.

This research links:

• Political economy / mobility literature (mentioned above), with

• economics-and psychology literature on strategic ignorance, overconfidence, self-deception, memory management, wishful thinking, etc.: Akerlof and Dickens (1982), Carrillo and Mariotti (2000), Bénabou and Tirole (2002), Mullainathan (2002), Weinberg (2000), Köszegi (2000) Landier (2001).

3 A simple model of ideological choice

There is a continuum of agents $i \in [0,1]$, who produce with the following technology:

$$y^i = \begin{cases} 1 & \text{with probability} \quad \pi^i + \theta e^i \\ 0 & \text{with probability} \quad 1 - (\pi^i + \theta e^i) \end{cases} \tag{1}$$

where e^i is their level of effort (alternatively, human capital resulting from an initial investment), θ is a parameter measuring the extent to which effort or acquired ability is rewarded, and π^i an innate or preexisting advantage –human or social capital inherited from one's parents, advantage due to discrimination, etc.. This variable takes values π_1 and π_0, for proportions $\varphi < 1/2$ and $1-\varphi$ of agents respectively; the average is $\bar{\pi} \equiv \varphi\pi_1 + (1-\varphi)\pi_0$. Similarly, we denote by $\bar{e}$ the average level of effort, and by $\bar{y} = \pi + \theta e$ the average output; both will be endogenous.

Income may be redistributed linearly at a tax rate $\tau \leq 1$, which will be determined through majority voting. Since there is a priori no reason to exclude regressive taxation, we allow $\tau < 0$. Imposing $\tau \in [0,1]$ would only (slightly) complicate the analysis.

Agents' preferences are subject, at the time the effort is exerted, to a "salience of present" effect measured by $1/\beta \geq 1$. Thus, the expected utility perceived by agent i when choosing e^i and facing a tax rate τ is:

$$U^i = E\left[(1-\tau)\left(y^i\right) + \tau\bar{y} - \frac{\left(e^i\right)^2}{2a\beta}\right] = E\left[(1-\tau)\left(\pi^i + \theta e^i\right) + \tau\left(\bar{\pi} + \theta\bar{e}\right) - \frac{\left(e^i\right)^2}{2a\beta}\right]. \tag{2}$$

Ex ante, however, he would evaluate the same payoff flows without the coefficient β. This means that the agent's ex post effort choice will always be suboptimally low, due to his lack of willpower ($\beta < 1$).

3.1 Signals and Beliefs

The true productivity of effort, θ, is unknown. We focus here on the more simple case where signals are perfectly correlated (reflecting for instance some aggregate information); the case where agents' signals are conditionally independent draws from a common distribution that depends on θ leads to similar results.[1] At the beginning of the period agents receive a common signal about the value of θ. With probability $1-q$ each one receives bad news, $\sigma = L$, and with probability q he receives no news at all, $\sigma = \emptyset$. In other words, "no news is good news"; see

[1] [Extensions to be added]. The case on which we focus here makes collective self-deception only more difficult, as initially all agents know the (relevant) true state of the world. On the other hand, by focusing on exogenous signals we are abstracting from the fact that the equilibrium tax rate τ may also reveal some information about θ. The more complex case where agents condition their effort decisions on this additional information leads to qualitatively similar results, however.

Figure 1. Let

$$\theta_L \equiv \mathsf{E}\,[\theta\,|\,\sigma = L] < \mathsf{E}\,[\theta\,|\,\sigma = \emptyset] \equiv \theta_H$$

be the expected values of the parameter θ conditional on these two possible events, and $\Delta\theta \equiv \theta_H - \theta_L$. For each agent i we denote his signal as $\sigma^i \in \{L, \varnothing\}$, and his information set just after receiving σ^i as Ω^i. When agents vote on taxes and choose effort levels, however, their information set is generally different, since they may not recollect certain signals received earlier (which could mean that either none was received, or that they have forgotten or repressed it). We denote their recollection and their information set at this later time as $\hat{\sigma}^i \in \{L, \varnothing\}$ and $\hat{\Omega}^i$ respectively. These define their posterior beliefs about θ, as well as about aggregate output, which depends on *other agents' beliefs* concerning θ :

$$\mu^i \equiv \Pr\left[\sigma = \varnothing \mid \hat{\Omega}^i\right], \tag{3}$$

$$\hat{\theta}^i \equiv E[\theta \mid \hat{\Omega}^i] = \mu^i\theta_H + (1-\mu^i)\theta_L \equiv \theta(\mu^i) \tag{4}$$

$$\hat{\Theta} \equiv E\left[\int_0^1 \hat{\theta}^j \, dj \,\middle|\, \hat{\Omega}^i\right]. \tag{5}$$

$$\hat{\Gamma}^i = E\left[\theta\cdot\hat{\Theta} \,\middle|\, \hat{\Omega}^i\right] \tag{6}$$

We now describe the mechanism through which agents may (partially) manipulate their own beliefs, or those of their children. As illustrated on Figure 1, let λ denote the probability that bad news (a signal $\sigma = L$) will later on be remembered accurately:

$$\lambda \equiv \Pr\left[\hat{\sigma} = L \mid \sigma = L\right] \tag{7}$$

We assume that an agent can increase or decrease this recall or awareness probability, at some cost $M(\lambda)$ that is minimized at a "natural" recall rate $\lambda_N \leq 1$.[2] Equivalently, one may think of an *intergenerational* mechanism for the transmission of beliefs and "values", with parents devoting time and resources $M(\lambda)$ to shielding or preserving their children's belief in a "just world", where effort is ultimately rewarded, in spite of evidence –perhaps the parents' own experience– that it may not be so just after all.

The optimal choice of λ, which is determined jointly with the political outcome (that is, through a general equilibrium mechanism), will be analyzed in Section 3.4. For the moment the only important features of the belief distortion mechanism are that: i) λ may be less than 1; ii) individuals are Bayesian (or at least, not completely naive) and therefore aware to some extent that they, and others, may have a systematic tendency to try and maintain a "belief in the just world". Consequently, they do not take absence of adverse recollections ($\hat{\sigma}^i = \varnothing$), or

[2] See Bénabou and Tirole (2002) for further discussions in light of the psychology literature.

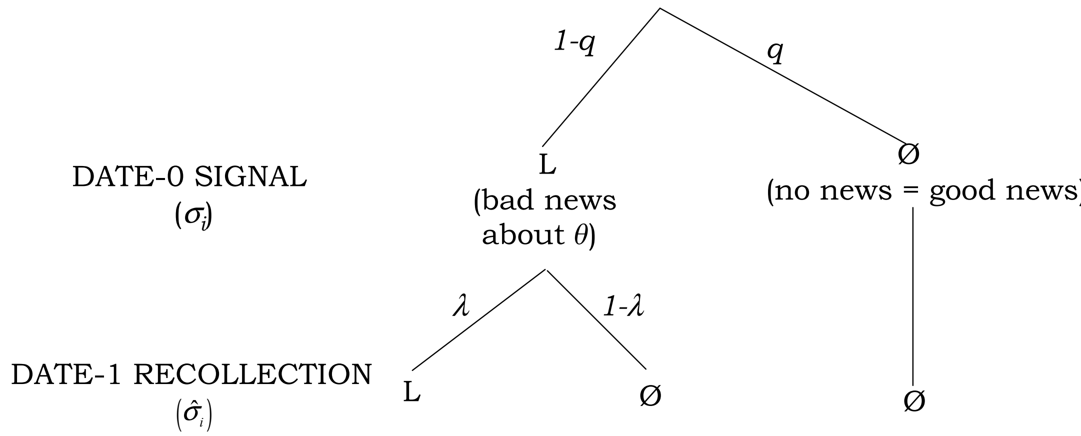

Figure 1: the manipulation of beliefs

his parents' exhortations that effort pays and crime does not, at face value. Instead, they assess the *reliability* of a "no bad news" recollection, $\hat{\sigma}^i = \varnothing$, as

$$r \equiv \Pr\left[\sigma = \emptyset \mid \hat{\sigma} = \emptyset; \lambda\right] = \frac{q}{q + (1-q)(1-\lambda)} \equiv r^*(\lambda). \tag{8}$$

This recall or information manipulation mechanism determines agents' *posterior beliefs* when voting and choosing effort: if $\hat{\sigma}^i = L$ then $\mu^i = 0$ and $\hat{\theta}^i = \theta_L$; if $\hat{\sigma}^i = \varnothing$ then $\mu^i = r$ and $\hat{\theta}^i = \theta(r)$. Finally, for all μ^i we can express (6) as

$$\hat{\Gamma}^i = \mu^i \theta_H \theta(r) + (1-\mu^i)\theta_L\left[\lambda\theta_L + (1-\lambda)\theta(r)\right] \equiv \Gamma(\mu^i; r, \lambda), \tag{9}$$

(or $\hat{\Gamma}(\mu^i)$ for short), where λ is the equilibrium strategy used by all agents (we will verify that everyone indeed chooses the same λ).[3]

3.2 Effort or investment decisions

Each agent i chooses effort optimally:

$$e^i = a\beta(1-\tau)\hat{\theta}^i, \tag{10}$$

implying that

$$E[\theta\bar{e} \mid \hat{\Omega}^i] = a\beta(1-\tau)\hat{\Gamma}^i. \tag{11}$$

[3] Thus when $\hat{\sigma}^i = L$ we have $\hat{\Gamma}^i = \Gamma(0; r, \lambda) = \theta_L^2 + (1-\lambda)\theta_L(\theta(r) - \theta_L)$, since $\mu^i = 0$.When $\hat{\sigma}^i = \varnothing$ we have $\hat{\Gamma}^i = \Gamma(r; r, \lambda) = r\theta_H\theta(r) + (1-r)\theta_L(\lambda\theta_L + (1-\lambda)\theta(r))$, since $\mu^i = r$. It will be useful to rewrite this last expression as: $\Gamma(r; r, \lambda) = \theta(r)\left[r\theta_H + (1-r)\theta_L\right] + (1-r)\theta_L\left[\lambda\theta_L - \lambda\theta(r)\right] = \theta(r)^2 - \lambda(1-r)\theta_L\left[\theta(r) - \theta_L\right].$

Note that it is not just agent i's beliefs concerning θ that are relevant for his political preferences but also *his beliefs about other agents' beliefs.* Indeed:

$$E[\bar{y} - \ \bar{\pi}|\ \hat{\Omega}^i] = E[\theta \bar{e} \mid \hat{\Omega}^i] = E\left[\theta \cdot \int_0^1 \ e^j \ dj \ \middle| \ \hat{\Omega}^i\right] \equiv a\beta(1-\tau)\hat{\Gamma}^i,$$

Given these beliefs, the agents' optimal choice of e^i results in an ex post utility (at the time the effort is chosen) of

$$U^i = (1-\tau)\left(\pi^i + a\beta(1-\tau)\left(\hat{\theta}^i\right)^2\right) + \tau(\bar{\pi} + a\beta(1-\tau)\hat{\Gamma}^i) - \frac{a\beta}{2}(1-\tau)^2\hat{\theta}^{i2}. \qquad (12)$$

Ex ante, however, the agent evaluates the same utility flow according to preferences that differ from U^i by the fact that the effort cost (represented by the last term) is no longer magnified by the salience parameter $1/\beta$. We shall capture both ex ante and ex post preferences by defining the function:

$$V^i \equiv (1-\tau)\left(\pi^i + a\beta(1-\tau)\left(\hat{\theta}^i\right)^2\right) + \tau(\bar{\pi} + a\beta(1-\tau)\hat{\Gamma}^i) - \frac{a\beta^2}{2\gamma}(1-\tau)^2\hat{\theta}^{i2}, \qquad (13)$$

where $\gamma = \beta$ corresponds to utility evaluated at the moment where effort in expended, while $\gamma = 1$ corresponds to utility before (and after) that time. This allows us in particular to cover both the case where agents vote over τ at the same time as they choose effort and that where they vote before choosing effort, and will then try to use τ to correct the underinvestment problem. Our results are robust to this modelling choice.

3.3 Social status, attitudes, and preferred tax rates

As intuition suggests, an agent's preferred tax rate decreases with the level of his "inherited" endowment, π. On the other hand, and somewhat surprisingly, it *need not* always *decrease with his degree of "optimism"* about the productivity of effort, θ (i.e., need not be lower when $\hat{\sigma} = \varnothing$ than when $\hat{\sigma} = L$). This is because a higher μ^i also raises the expected level of aggregate output, from which transfers are funded. We shall therefore have to look for conditions that ensure that this "tax base" effect is dominated by the "own income" (net of effort) effect.

Assumption 1 *Assume that:*

$$\begin{aligned} \frac{\Delta\theta}{\theta_L} &< \frac{2\beta}{\gamma}, \\ \left(1 - \frac{\beta}{2\gamma}\right)\theta_L^2 &< \frac{\bar{\pi} - \pi_0}{\beta a} < \theta_L^2. \end{aligned}$$

For all $(\pi, \mu; \lambda, r)$, we shall denote as

$$T(\pi, \mu; \lambda, r) \equiv 1 - \frac{\pi - \bar{\pi} + a\beta\Gamma(\mu)}{a\beta\left[2\Gamma(\mu) - (2 - \beta/\gamma)\,\theta(\mu)^2\right]}, \tag{14}$$

the solution to the first-order conditions $\partial V^i/\partial\tau = 0$, for any $\pi^i = \pi$ and $\mu^i = \mu$. (Note that $T(\pi, \mu; \lambda, r)$ need not a priori be less than 1). With a slight abuse of notation we shall denote in particular

$$\begin{aligned} T_L(\pi; \lambda, r) &\equiv T(\pi, 0; \lambda, r), \\ T_\varnothing(\pi; \lambda, r) &\equiv T(\pi, r; \lambda, r), \end{aligned}$$

the values of these functions corresponding to the two posteriors $\mu = 0$ and $\mu = r$ that agent with recollections $\hat{\sigma} = L$ and $\hat{\sigma} = \varnothing$ can have in an equilibrium. We shall refer to such agents as, respectively, "optimists" and "pessimists".

Proposition 1 *Under Assumption 1, each agent's preferences are strictly concave in τ, and his preferred tax rate τ^i is $T_L(\pi^i; \lambda, r)$ when he recalls an adverse signal ($\hat{\sigma}^i = L$), and $T_\varnothing(\pi^i; \lambda, r)$ when he does not ($\hat{\sigma}^i = \varnothing$). These preferred tax rates are always decreasing in the individual's initial endowment π^i, and furthermore:*

$$\begin{aligned} T_\varnothing(\pi_0; \lambda, r) &< T_L(\pi_0; \lambda, r) < 1 \\ T_\varnothing(\pi_1; \lambda, r) &< 0 < T_L(\pi_0; \lambda, r) \end{aligned}$$

Consider now the state of the world where the news about θ that agents initially receive are bad: $\hat{\sigma}^I = L$. At the time of voting, the results in Proposition 1 show that the pessimistic poor (i.e., those who recall the bad news) always want the highest tax rate $T_L(\pi_0; \lambda, r)$. If the equilibrium degree of recall λ is high enough that $(1 - \varphi)\lambda > 1/2$, they will be a majority, and impose their choice of policy. When the degree of forgetting or repression is high enough that $(1 - \varphi)\lambda < 1/2$, on the other hand, they will be a minority, and moreover since they have the most "extreme preferences" (the highest desired tax rate) they will also not be pivotal. Two cases may then occur, illustrated on the upper and lower panels of Figure 2:

Case 1: if $T_L(\pi_1; \lambda, r) < T_\varnothing(\pi_0; \lambda, r)$, then

$$\max\left\{T_L(\pi_1; \lambda, r), T_\varnothing(\pi_1; \lambda, r)\right\} < T_\varnothing(\pi_0; \lambda, r) < T_L(\pi_0; \lambda, r), \tag{15}$$

and since the poor overall are a majority, the pivotal group is now that of the optimistic poor, which sets the tax rate $T_\varnothing(\pi_0; \lambda, r)$.

Case 2: if $T_L(\pi_1; \lambda, r) > T_\varnothing(\pi_0; \lambda, r)$, then

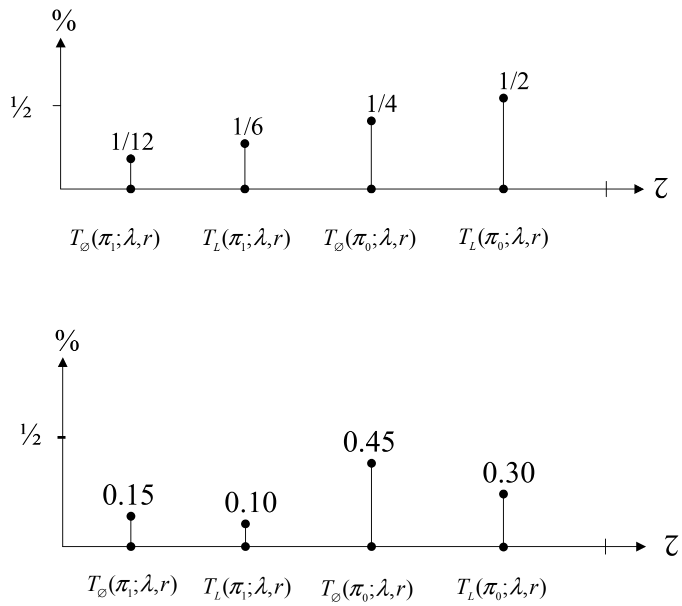

Figure 2: examples of low-BJW (upper panel, $\lambda = 2/3$) and high-BJW (lower panel, $\lambda = 1/4$) dominant ideologies, for $\varphi = 1/4$. In the first case the pivotal agent is disadvantaged and pessimistic, and sets $\tau = T_L(\pi_0)$. In the second case he is disadvantaged but optimistic, and sets $\tau = T_\varnothing(\pi_0)$.

$$T_\varnothing(\pi_1;\lambda,r) < T_\varnothing(\pi_0;\lambda,r) < T_L(\pi_1;\lambda,r) < T_L(\pi_0;\lambda,r). \tag{16}$$

Therefore if $\lambda < 1/2$ the optimists (rich plus poor) constitute a majority, so the pivotal group is again the optimistic poor, and the tax rate $T_\varnothing(\pi_0;\lambda,r)$. If $\lambda > 1/2$, on the other hand, the pivotal group is that of the pessimistic rich, who set the tax rate $T_L(\pi_1;\lambda,r)$.

In summary, for the pivotal vote to switch from the pessimistic poor to a group that desires a lower tax rate, therefore, it must be that the equilibrium recall probability decline from a value such that $(1-\varphi)\lambda > 1/2$ to a value such that $(1-\varphi)\lambda' < 1/2$.

Of course this recall probability is endogenous, resulting from agents' repression or rehearsal decisions, which themselves depend on the taxes and transfers that they anticipate will prevail at the time of effort. We therefore now turn to the determination of these motivated beliefs, and to the fixed-point problem that ultimately defines an equilibrium.

3.4 Memory and repression

Consider now agent i's expected utility at the start of the period, i.e. at the time he receives his signal σ^i. This expected utility, denoted $\tilde{U}^i$, differs from U^i (utility perceived at the time of effort), for two reasons. First, the effort cost is not subject to a salience-of-the present effect.

Second, the agent's information set at this point, Ω^i, includes the knowledge of the actual signal $\sigma^i \in \{L, \varnothing\}$ that he has received. By contrast, when he votes and chooses effort later on, his decisions will be based on the information set $\hat{\Omega}^i$, in which σ^i has been replaced by its (less informative, or "garbled") subjective recollection $\hat{\sigma}^i \in \{L, \varnothing\}$. Thus:

$$\begin{aligned}\tilde{U}^i &\equiv E\left[(1-\tau)y^i + \tau\bar{y} - \frac{(e^i)^2}{2a} \mid \Omega^i\right] = (1-\tau)\pi^i + \tau\bar{\pi} + a\beta\tau(1-\tau)E[\theta \cdot \Theta \mid \Omega^i] \\ &\quad + a\beta(1-\tau)^2\ E\left[E[\theta \mid \hat{\Omega}^i] \cdot \left(E[\theta \mid \Omega^i] - \frac{\beta}{2}\ E[\theta \mid \hat{\Omega}^i]\right) \Big| \ \Omega^i\right]\ , \end{aligned} \tag{17}$$

where τ is the tax rate that he anticipates will be chosen by society.

When $\sigma^i = \varnothing$, the agent has no decision to take with respect to memory. Let us therefore focus on the case where $\sigma^i = L$. If he ends up with posterior belief μ, the agent will exert effort $e^i = \beta a(1-\tau)\left(\mu\theta_H + (1-\mu)\theta_L\right)$, and thus achieve the utility level

$$\begin{aligned}\tilde{U}_L(\pi, \tau, \mu; \lambda, r) &\equiv (1-\tau)\pi^i + \tau\bar{\pi} + a\beta\tau(1-\tau)\theta_L\left[\lambda\theta_L + (1-\lambda)\theta(r)\right] \\ &\quad + a\beta(1-\tau)^2\left(\mu\theta_H + (1-\mu)\theta_L\right)\left[\theta_L - \frac{\beta}{2}\left(\mu\theta_H + (1-\mu)\theta_L\right)\right]. \end{aligned} \tag{18}$$

Note here that (in contrast to what happened in V^i), μ^i does not affect the size of the transfer which the agent expects to receive (second term in (18)). His expectation of what aggregate income will ultimately turn out to be reflects his current information (in this instance, $E\left[\theta \mid \sigma^i = L\right] = \theta_L$), not the possibly distorted recollections of that data that he may have later on.[4]

An individual who recalls $\hat{\sigma}^i = L$ will have $\mu^i = 0$, whereas for $\hat{\sigma}^i = \varnothing$ he will have $\mu^i = r$, where (r, λ) denotes the (symmetric) equilibrium strategy played by all agents.[5] The cognitive optimization problem for an agent who receives the signal $\sigma^i = L$ is therefore:

$$\max_{\lambda' \in [0,1]} \left\{\lambda' \tilde{U}_L(\pi, \tau, 0; \lambda, r) + (1-\lambda')\tilde{U}_L(\pi, \tau, r; \lambda, r) - M(\lambda')\right\}, \tag{19}$$

where $M(\lambda')$ is the cost of achieving a recall (or intergenerational transmission) probability equal to λ. A typical cost function is represented by the U-shaped curve on Figure 3, where $\bar{\lambda}$ represents the natural rate of recall.

Given (18), we can rewrite the optimal-awareness problem as:

[4] Things would be different if the agent at date 0 cared not just about expected final payoffs, but also derived "anticipal utility" from the interim level of utility achieved at $t = 1$. In that case there would be consumption value to holding optimistic views about the size of aggregate output (which depends on θ), because it would allow the agent to (temporarily) savor the prospects of receiving a large transfer.

[5] The fact that π^i does not interact with beliefs in this expression makes clear that the optimal cognitive strategy is independent of initial endowments.

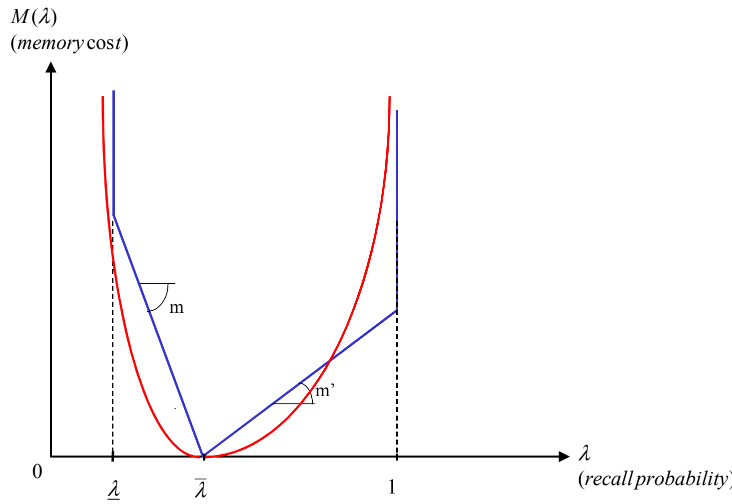

Figure 3: the awareness technology

$$\max_{\lambda' \in [0,1]} \left\{ \beta a (1-\tau)^2 \left[\lambda' \left(1 - \frac{\beta}{2}\right) \theta_L + (1-\lambda') \left(1 - \frac{\beta\theta(r)}{2\theta_L}\right) \theta(r) \right] - M(\lambda') \right\}. \tag{20}$$

Two key effects are apparent in this formula:

• *Role of time inconsistency:* let $M \equiv 0$. When $\beta \approx 1$, agents will choose $\lambda' = 1$ (information is always valuable); when $\beta \approx 0$, they always choose $\lambda' = 0$ (self-motivation is critical).

• *Role of taxes:* assume that β is low enough that repression is valuable, but now also costly ($M' > 0$). Then, the lower is τ, the greater is the incentive to repress, that is, to choose low λ'. This is the second *complementarity* mechanism discussed earlier.[6]

To simplify the problem, we shall take the memory-cost function to be piecewise linear, with natural (costless) rate of recall $\bar{\lambda} \in (0, 1]$, a minimum rate of recall $\underline{\lambda} \in [0, \bar{\lambda})$ (or maximum degree of repression $1 - \underline{\lambda} > 1 - \bar{\lambda}$), and linear marginal costs $m > 0$ and $m' > 0$ for repression and rehearsal respectively (see Figure 3).

Assumption 2 *The memory cost function is given by:*

$$M(\lambda) = \begin{cases} +\infty & \textit{for} \quad \lambda < \underline{\lambda} \\ m(\bar{\lambda} - \lambda) & \textit{for} \quad \lambda \in [\underline{\lambda}, \bar{\lambda}] \\ m'(\lambda - \bar{\lambda}) & \textit{for} \quad \lambda \geq \bar{\lambda} \end{cases} .$$

[6] [But a bit more complicated: in equilibrium, lower τ also goes with lower (λ, r), hence higher $\theta(r)$].

3.5 Politico-ideological equilibria

We are now able to characterize a (symmetric) politico-economic equilibrium as triplet (λ, r, τ) such that:

$$\lambda \in \arg\max_{\lambda' \in [0,1]} \left\{ \lambda' \tilde{U}_L(\pi, \tau, 0; \lambda, r) + (1-\lambda')\tilde{U}_L(\pi, \tau, r; \lambda, r) - M(\lambda') \right\}, \tag{21}$$

$$r = \frac{q}{q + (1-q)(1-\lambda)}, \tag{22}$$

τ is the majority tax rate, given the distribution of beliefs induced by (λ, r). (23)

We shall specifically look for two equilibria, defined by $(\underline{\lambda}, \underline{r}, \underline{\tau})$ and $(\bar{\lambda}, \bar{r}, \bar{\tau})$ such that:

1) When agents do not repress bad news about θ very much $(\lambda = \bar{\lambda})$, enough of the poor end up with (correctly) pessimistic beliefs $\mu^i = 0$ that they constitute a majority, and thus impose a high tax rate $\bar{\tau} = T_L(\pi_0; \bar{\lambda}, \bar{r})$. This requires that $(1-\varphi)\bar{\lambda} > 1/2$. The expectation of a high tax rate, and therefore a low return to effort, generates in turn only weak incentives to repress the fact that θ is low. So individuals indeed make no effort at repression, choosing the natural recall rate $\bar{\lambda}$;

2) When agents try hard to repress bad news about θ $(\lambda = \underline{\lambda})$ enough of the poor end up with relatively optimistic beliefs $\mu^i = \bar{r}$ that $(1-\varphi)\underline{\lambda} < 1/2$. As explained earlier, this implies that either:

a) the optimistic poor constitute a pivotal minority that gets to impose its preferred tax rate, $\underline{\tau} = T_\varnothing(\pi_0; \underline{\lambda}, \underline{r}) \ < T_L(\pi_0; \underline{\lambda}, \underline{r})$;

b) the pivotal group is the optimistic rich, and the pessimistic poor side with them to impose the tax rate $\underline{\tau} = T_\varnothing(\pi_1; \underline{\lambda}, \underline{r}) \in \left(T_\varnothing(\pi_0; \underline{\lambda}, \underline{r}), T_L(\pi_0; \underline{\lambda}, \underline{r})\right)$; this requires $\underline{\lambda} > 1/2$.

In both cases, the expectation of a relatively low tax rate, and therefore a high return to effort, generates in turn strong incentives to repress the fact that θ is low. So individuals indeed make significant efforts at repression, which implies that a high fraction $1 - \bar{\lambda}$ of them do forget the adverse information.[7]

The following assumption which ensures that the pivotal group switches from the pessimistic poor to a group that desires a lower tax rate (either the optimistic poor or the pessimistic rich) as λ declines from $\underline{\lambda}$ to $\bar{\lambda}$.

Assumption 3 $(1-\varphi)\underline{\lambda} < 1/2 < (1-\varphi)\bar{\lambda}$.

[7] In addition to these extremal equilibria, there may also be an equilibrium (or equilibria) where the first-order condition with respect to λ' holds with equality at $\lambda \in (\underline{\lambda}, \bar{\lambda})$.

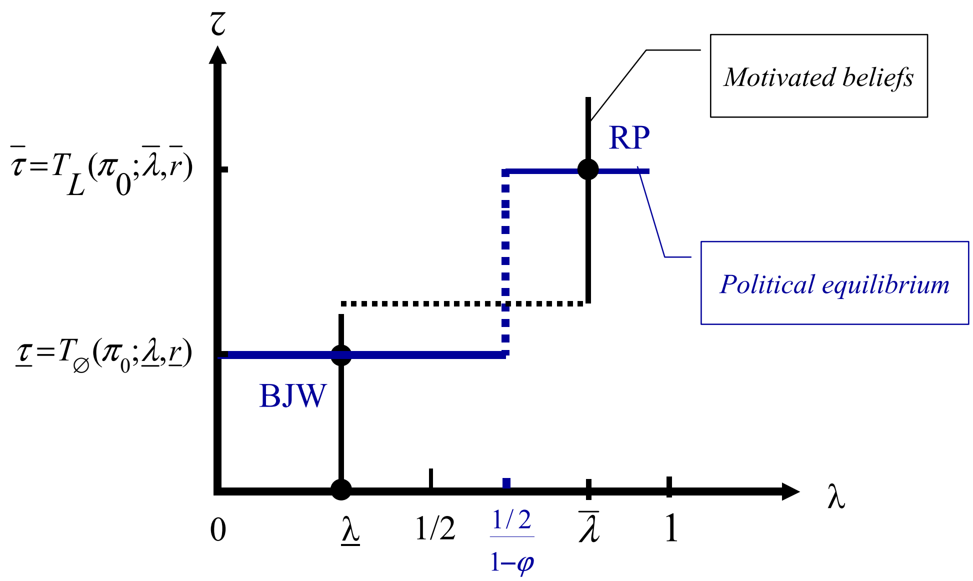

Figure 4: Ideological choices, political choices, and the set of equilibria (BJW: Belief in a Just World; RP: Realistic Pessimism).

The requirements for the two politico-economic equilibria are:

$$\begin{cases} \tilde{U}_L(\pi,\bar{\tau},\bar{r};\bar{\lambda},\bar{r}) - \tilde{U}_L(\pi,\bar{\tau},0;\bar{\lambda},\bar{r}) & < & m\left(\bar{\lambda}-\underline{\lambda}\right) \\ \tilde{U}_L(\pi,\underline{\tau},\underline{r};\underline{\lambda},\underline{r}) - \tilde{U}_L(\pi,\underline{\tau},0;\underline{\lambda},\underline{r}) & > & m\left(\bar{\lambda}-\underline{\lambda}\right) \end{cases} \tag{24}$$

$$\begin{cases} \tilde{U}_L(\pi,\bar{\tau},\bar{r};\bar{\lambda},\bar{r}) - \tilde{U}_L(\pi,\bar{\tau},0;\bar{\lambda},\bar{r}) & > & -m'(1-\bar{\lambda}) \\ \tilde{U}_L(\pi,\underline{\tau},\underline{r};\underline{\lambda},\underline{r}) - \tilde{U}_L(\pi,\underline{\tau},0;\underline{\lambda},\underline{r}) & > & -\,m'(1-\underline{\lambda}) \end{cases} \tag{25}$$

for all π,[8] together with

To establish (by construction) the existence of these two equilibria, let us start with $\underline{\lambda}$ and $\bar{\lambda}$ given by the memory technology, and satisfying Assumptions (2)–(3). We then define $\underline{r}$ and $\bar{r}$ from Bayes' rule (21), $\theta(\underline{r})$ and $\theta\ (\bar{r})$ in the usual way, and use (14) to compute

$$\bar{\tau} \equiv T_L(\pi_0;\bar{\lambda},\bar{r}), \tag{26}$$

$$\underline{\tau} \equiv \begin{cases} T_\varnothing(\pi_0;\underline{\lambda},\underline{r}) & \text{if} \quad \underline{\lambda} \leq 1/2 \\ \max\left\{T_\varnothing(\pi_0;\underline{\lambda},\underline{r}), T_L(\pi_1;\underline{\lambda},\underline{r})\right\} & \text{if} \quad \underline{\lambda} > 1/2 \end{cases}. \tag{27}$$

A first issue is whether it is indeed the case that $\underline{\tau} < \bar{\tau}$. This is in fact not obvious, since the knowledge that other agents are likely to be more optimistic (due to their using the recall strategy $\underline{\lambda}$ rather than $\bar{\lambda}$), and therefore to work harder, tends to make a poor individual want

[8] It is easily seen that the differences in (24)–(25) are in fact independent of π.

to tax them more. We shall need again the conditions in Assumption 1 that ensure that this tax base effect (now operating through *other* agents' beliefs) is dominated by the direct concern for one's own income (net of effort costs).

Proposition 2 *Under Assumption 1, the tax rate $\underline{\tau}$ is less than $\bar{\tau} \equiv T_L(\pi_0; \bar{\lambda}, \bar{r})$.*

Finally, we verify that any individual's incentive to forget or repress bad news about θ is indeed higher in a low-tax, high-repression politico-economic environment ($(\lambda, r, \tau) = (\underline{\lambda}, \underline{\tau}, \underline{r})$) than in a high-tax, high-repression environment (and strictly positive in the first case):

$$\max\left\{\tilde{U}(\pi, \bar{\tau}, \bar{r}; \bar{\lambda}, \bar{r}) - \tilde{U}(\pi, \bar{\tau}, 0; \bar{\lambda}, \bar{r}), 0\right\} < \tilde{U}(\pi, \underline{\tau}, \underline{r}; \underline{\lambda}, \underline{r}) - \tilde{U}(\pi, \underline{\tau}, 0; \underline{\lambda}, \underline{r}), \tag{28}$$

for all π. Under (28), the fixed-point conditions for $\underline{\lambda}$ and $\bar{\lambda}$ given by (24) will indeed hold for all $m > 0$ in the range between these two terms divided by $\bar{\lambda} - \underline{\lambda}$. We show in the appendix that the following are sufficient conditions for (28) to hold, and also to ensure (when $\bar{\lambda} < 1$) that no agent want to rehearse the bad news.

Assumption 4 *Let*

$$1 + \left(\frac{\Delta\theta}{2\theta_L}\right)\underline{r} < \frac{1}{\beta} < 1 + \left(\frac{\Delta\theta}{2\theta_L}\right)(1 + \underline{r}),$$

if $\bar{\lambda} = 1$, or

$$1 + \left(\frac{\Delta\theta}{2\theta_L}\right)\bar{r} < \frac{1}{\beta} < 1 + \left(\frac{\Delta\theta}{2\theta_L}\right)(1 + \underline{r}),$$

if $\bar{\lambda} < 1$. [Probably want to focus on one of these two cases].

This yields our main result, illustrated on Figure 4.

Proposition 3 *Assume that Assumptions 1–4 are satisfied. Then, for a range of values of the repression cost m (and for all $m' > 0$), there exist two politico-economic equilibria, with degrees of repression $\underline{\lambda}$ and $\bar{\lambda}$ and associated tax rates $\underline{\tau}$ and $\bar{\tau}$, such that $\underline{\lambda} < \bar{\lambda}$ and $\underline{\tau} < \bar{\tau}$.*

Some Properties of the Equilibria:

• In the bad state of the world (not-so-just): tax rate is lower, and aggregate output is higher, in BJW equilibrium than in welfare state equilibrium.

• In good state of the world (very just): rankings ambiguous, due to Bayesian "self-doubt effect". Could eliminate / attenuate.

4 Extensions of the basic model

4.1 Simple intergenerational dynamics

• Suppose that initial advantage, π, simply reflects parental resources (human of financial wealth). Accumulates. Then,

$$\varphi_{t+1} = \varphi_t \pi_1 + (1-\varphi_t)\pi_0 + \beta a(1-\tau_t)\theta_L \left[(\lambda\theta_L + (1-\lambda)\theta(r_t))\right]$$

$\Rightarrow$ in steady-state (not-so just world):

$$1-\varphi = \frac{\beta a(1-\tau)\theta_L \left[(\lambda\theta_L + (1-\lambda)\theta(r))\right]}{\pi_1 - \pi_0}.$$

• Amplifying mechanism.

4.2 The lazy poor

• Suppose that a fraction x of people are lazy / have no willpower: $a = 0$, or $\beta = 0$ (latter actually more relevant) $\Rightarrow$ never work ($e = 0$), no incentive to maintain BJW ($\lambda = \bar{\lambda}$).
• "Laziness" and initial endowment $\pi \in \{\pi_0, \pi_1\}$ uncorrelated, for simplicity. Assume x is small enough that lazy people are never pivotal (or, too lazy to vote?).
• When one sees a person who has failed in life / is poor ex-post ($y = 0$), what is the probability that it is due to laziness? For an agent i :

$$p \equiv \Pr\left[\beta = 0 \mid y = 0, \hat{\Omega}^i\right] = \frac{(1-\bar{\pi})x}{(1-\bar{\pi})x + (1-x)(1-\bar{\pi} - a\beta(1-\tau)\hat{\Gamma}^i)}$$

Since (in state θ_L) :

• $(1-\tau)$ is higher in BJW equilibrium (high denial $1-\underline{\lambda}$) than in realistic-pessimism (RP) equilibrium

• The politically pivotal voter $\hat{\Gamma}^i$ often also has higher estimate of the average contribution of equilibrium effort to success in BJW equilibrium (pivotal = pessimistic poor, under some conditions) than in RP equilibrium (pivotal = pessimistic poor).

$\Rightarrow$ there is greater "stigma" on the (ex-post) poor in a BJW equilibrium.

• Consequences of these negative inferences / stigma stereotypes: emotional (resentment, anger, etc.) and / or economic: want to transfer less to them. Selective altruism, where people want to help the non-lazy poor only.

• Also: people do not know the fraction x of the poor who are lazy. Make inferences about it from: i) observed poverty rate; ii) their own beliefs about θ. Again, if pivotal group has BJW $\Rightarrow$ think more of the poor are lazy $\Rightarrow$ want to give less transfers to those who failed $\Rightarrow$ will require lower taxes to finance them $\Rightarrow$ greater incentives to hold BJW.

Appendix

Proof of Proposition 1

1) Proof of concavity: we have

$$\begin{aligned}
\frac{\partial V^i}{\partial \tau} &\equiv \bar{\pi} - \pi^i + (1-2\tau)a\beta\hat{\Gamma}^i - \left(2 - \frac{\beta}{\gamma}\right) a\beta(1-\tau)\left(\hat{\theta}^i\right)^2, \\
\frac{\partial^2 V^i}{\partial \tau^2} &= a\beta\left[(2-\beta/\gamma)\left(\hat{\theta}^i\right)^2 - 2\hat{\Gamma}^i\right]
\end{aligned}$$

The function V^i is concave in τ if $(2-\beta/\gamma)\left(\hat{\theta}^i\right)^2 < 2\hat{\Gamma}^i$, meaning that:

$$(2-\beta/\gamma)\left(\mu^i\theta_H + (1-\mu^i)\theta_L\right)^2 < 2\left[\mu^i\theta_H\theta(r) + (1-\mu^i)\theta_L\left(\lambda\theta_L + (1-\lambda)\theta(r)\right)\right].$$

Since the difference between the left- and right-hand sides is quadratic and convex in μ^i, it only needs to be checked at the boundaries of the range of beliefs $[0, r]$ achievable in equilibrium. For $\mu^i = 0$ we get $(2-\beta/\gamma)(\theta_L)^2 < 2\theta_L[\lambda\theta_L + (1-\lambda)\theta(r)]$, which trivially holds. For $\mu^i = r$, we require that:

$$\begin{aligned}
(2-\beta/\gamma)\,\theta(r)^2 &\leq 2\left[r\theta_H\theta(r) + (1-r)\theta_L(\lambda\theta_L + (1-\lambda)\theta(r))\right] \\
&= 2\left[(r\theta_H + (1-r)\theta_L)\,\theta(r) - (1-r)(1-\lambda)\theta_L(\theta(r)-\theta_L)\right] \\
&= 2\left[\theta(r)^2 - (1-r)(1-\lambda)\theta_L(\theta(r)-\theta_L)\right] \iff \\
2(1-r)(1-\lambda)\theta_L(\theta(r)-\theta_L) &\leq (\beta/\gamma)\,\theta(r)^2 \iff \\
2r(1-r)(1-\lambda)\theta_L(\theta_H-\theta_L) &\leq (\beta/\gamma)\left(r\theta_H + (1-r)\theta_L\right)^2,
\end{aligned}$$

Since $r(1-r) \leq 1/4$, it is sufficient that:

$$\theta_H - \theta_L \leq 2\,(\beta/\gamma)\,\theta_L, \tag{A.1}$$

which is ensured by the first part of Assumption (1). We now examine how agents' preferred tax rates rank, as functions of their endowments and beliefs. ||

2) Proof that $T_\varnothing(\pi_0;\lambda,r) < T_L(\pi_0;\lambda,r)$: For any π, $T_\varnothing(\pi;\lambda,r) < T_L(\pi;\lambda,r)$ if and only if $1 - T_L(\pi;\lambda,r) < 1 - T_\varnothing(\pi,r;\lambda,r)$, or:

$$\frac{\pi - \bar{\pi} + a\beta\Gamma(0)}{a\beta\left[2\Gamma(0) - (2-\beta/\gamma)\,\theta(0)^2\right]} < \frac{\pi - \bar{\pi} + a\beta\Gamma(r)}{a\beta\left[2\Gamma(r) - (2-\beta/\gamma)\,\theta(r)^2\right]}, \tag{A.2}$$

which is equivalent to:

$$\left(\frac{\bar{\pi}-\pi}{a\beta}\right)\left[\left(2-\frac{\beta}{\gamma}\right)(\theta(r)^2-\theta_L^2)-2\left(\Gamma(r)-\Gamma(0)\right)\right]<\left(2-\frac{\beta}{\gamma}\right)\left[\theta(r)^2\Gamma(0)-\theta(0)^2\Gamma(r)\right]. \tag{A.3}$$

Now, note that:

$$\begin{aligned}\Gamma(r)-\Gamma(0) &= \theta(r)^2-\theta_L^2-\lambda(1-r)\theta_L(\theta(r)-\theta_L)-(1-\lambda)\theta_L(\theta(r)-\theta_L)\\ &= r\,(\Delta\theta)\left[\theta(r)+\theta_L-(1-\lambda r)\theta_L\right]=r\,(\Delta\theta)\left[(1+\lambda r)\theta_L+r\,(\Delta\theta)\right] \qquad \text{(A.4)}\end{aligned}$$

and that:

$$\begin{aligned}\theta(r)^2\Gamma(0)-\theta(0)^2\Gamma(r) &= \theta_L\theta(r)^2\left[\theta_L+(1-\lambda)(\theta(r)-\theta_L)\right]-\theta_L^2\left[\theta(r)^2-\lambda(1-r)\theta_L\left(\theta(r)-\theta_L\right)\right]\\ &= \theta_L(\theta(r)-\theta_L)\left[(1-\lambda)\theta(r)^2+\lambda(1-r)\theta_L^2\right].\\ &= r\theta_L\left(\Delta\theta\right)\left[(1-\lambda)\theta(r)^2+\lambda(1-r)\theta_L^2\right] \qquad \text{(A.5)}\end{aligned}$$

Therefore, condition (A.2) takes the form:

$$\begin{aligned}\left(2-\frac{\beta}{\gamma}\right)\theta_L\left[(1-\lambda)\theta(r)^2+\lambda(1-r)\theta_L^2\right] &> \left[\left(2-\frac{\beta}{\gamma}\right)(\theta(r)+\theta_L)-2\left(1+\lambda r\right)\theta_L-2r\left(\Delta\theta\right)\right]\\ &\quad\times\left(\frac{\bar{\pi}-\pi}{a\beta}\right)\\ &= \left[\left(2-\frac{\beta}{\gamma}\right)(2\theta_L+r\left(\Delta\theta\right))-2\left(1+\lambda r\right)\theta_L-2r\left(\Delta\theta\right)\right]\\ &\quad\times\left(\frac{\bar{\pi}-\pi}{a\beta}\right)\\ &= \left[2\left(1-\lambda r-\frac{\beta}{\gamma}\right)\theta_L-\left(\frac{\beta}{\gamma}\right)r\left(\Delta\theta\right)\right]\left(\frac{\bar{\pi}-\pi}{a\beta}\right). \quad \text{(A.6)}\end{aligned}$$

If the term in brackets on the right-hand side is negative –this always occurs, in particular, when $\gamma=\beta$– the condition automatically holds for the poor, since for them $\bar{\pi}-\pi_1>0$. When the right-hand side is positive (this only occurs when $\gamma=1$) the condition always holds for the rich ($\bar{\pi}-\pi_0<0$), implying that $T_{\varnothing}(\pi_1;\lambda,r)<T_L(\pi_1;\lambda,r)$. The claim to be shown, however, pertains to the poor. In order to show that

$$F(r,\lambda)\equiv\left(2-\frac{\beta}{\gamma}\right)\theta_L\left[(1-\lambda)\theta(r)^2+\lambda(1-r)\theta_L^2\right]-\left[2\left(1-\lambda r-\frac{\beta}{\gamma}\right)\theta_L-\left(\frac{\beta}{\gamma}\right)r\left(\Delta\theta\right)\right]\left(\frac{\bar{\pi}-\pi_0}{a\beta}\right),$$

is always positive, let us first observe that, for given λ, this is a convex, quadratic function in r, with:

$$\begin{aligned}\frac{\partial F(r,\lambda)}{\partial r} &= \left(2-\frac{\beta}{\gamma}\right)\theta_L\left[2(1-\lambda)\theta(r)r\left(\Delta\theta\right)-\lambda\theta_L^2\right]+\left[2\lambda\theta_L+\left(\frac{\beta}{\gamma}\right)r\left(\Delta\theta\right)\right]\left(\frac{\bar{\pi}-\pi_0}{a\beta}\right)\\ &> 2\lambda\theta_L\left[\left(\frac{\bar{\pi}-\pi_0}{a\beta}\right)-\left(1-\frac{\beta}{2\gamma}\right)\theta_L^2\right]>0,\end{aligned}$$

by the first inequality in the second condition of Assumption 1. Therefore $F(r,\lambda)>0$ for ail $r\in[0,1]$ if and only if: $F(0,\lambda)>0$, which is equivalent to

$$\left(1-\frac{\beta}{2\gamma}\right)\theta_L^2>\left(1-\frac{\beta}{\gamma}\right)\left(\frac{\bar{\pi}-\pi_0}{a\beta}\right), \tag{A.7}$$

Since $(1-\beta/2\gamma)/(1-\beta/\gamma)>1$, this inequality is ensured by the second inequality in the second condition of Assumption 1. $\|$

3) Proof that $T_L(\pi_0;\lambda,r)>0:$ this is equivalent to $1-T(\pi_0,0;\lambda,r)<1$, or by (14):

$$\begin{aligned}\frac{\pi_0-\bar{\pi}}{a\beta}+\Gamma(0) &< 2\Gamma(0)-\left(2-\beta/\gamma\right)\theta(0)^2 \iff\\ \frac{\bar{\pi}-\pi_0}{a\beta} &> \theta_L\left[\left(2-\beta/\gamma\right)\theta_L-\left(\theta_L+(1-\lambda)r\left(\Delta\theta\right)\right)\right]\iff\\ \frac{\bar{\pi}-\pi_0}{a\beta} &> \theta_L\left[\left(1-\beta/\gamma\right)\theta_L-(1-\lambda)r\left(\Delta\theta\right)\right].\end{aligned}$$

A sufficient condition is that:

$$\frac{\bar{\pi}-\pi_0}{a\beta}>\left(1-\frac{\beta}{\gamma}\right)\theta_L^2.$$

It is automatically satisfied when $\gamma=\beta$, and in any case is ensured by the second condition in Assumption 1. $\|$

4) Proof that $T_L(\pi_0;\lambda,r)<1:$ this is equivalent to $1-T(\pi_0,0;\lambda,r)>0$, or by (14):

$$\frac{\bar{\pi}-\pi_0}{\beta a}<\Gamma(0)=\theta_L\left[(1-\lambda)\theta(r)+\lambda\theta_L\right],$$

for which it is sufficient that $\bar{\pi}-\pi_0<\beta a\theta_L^2$, which is ensured by the second condition in Assumption 1. $\|$

5) Proof that $T_\varnothing(\pi_1;\lambda,r)<0:$ by (14), this is equivalent to:

$$\frac{\pi_1-\bar{\pi}}{a\beta}>\Gamma(r)-\left(2-\beta/\gamma\right)\theta(r)^2=\Gamma(r)-\theta(r)^2-\left(1+\beta/\gamma\right)\theta(r)^2,$$

which holds automatically since $\theta(r)^2>\Gamma(r)$.$\|$

6) Proof that agents i's preferred tax rates is $T_L(\pi^i;\lambda,r)$ or $T_\varnothing(\pi^i;\lambda,r)$, depending on $\hat{\sigma}^i=L,\varnothing:$

by concavity of the objective function, we have:[9]

$$\tau^i = \min\left\{T(\pi, \mu^i; \lambda, r), 1\right\}.$$

Furthermore, we have established that:

$$\begin{aligned} T_{\varnothing}(\pi_1; \lambda, r) &< T_{\varnothing}(\pi_0; \lambda, r) < T_L(\pi_0; \lambda, r) < 1 \quad &(A.8)\\ \max\{T_L(\pi_1; \lambda, r), 0\} &< T_L(\pi_0; \lambda, r), \quad &(A.9)\end{aligned}$$

so $T_L(\pi_0; \lambda, r)$ is the largest desired tax rate, and the constraint $\tau \leq 1$ is never binding in equilibrium. ■

Proof of Proposition 2

Since $T_L(\pi_1; \underline{\lambda}, \underline{r}) < T_L(\pi_0; \underline{\lambda}, \underline{r})$ and $T_{\varnothing}(\pi_0; \underline{\lambda}, \underline{r}) < T_L(\pi_0; \underline{\lambda}, \underline{r})$ by Proposition 1, it will be sufficient for $\underline{\tau} < \bar{\tau}$ that $T_L(\pi_0; \lambda, r)$ be increasing in (λ, r) for all (λ, r) satisfying (21) or, equivalently, that

$$\begin{aligned} 1 - T_L(\pi_0; \lambda, r) &\equiv \frac{\pi_0 - \bar{\pi} + \beta a \theta_L\left[\theta_L + (1-\lambda) r(\theta_H - \theta_L)\right]}{\beta a \theta_L\left[(\beta/\gamma)\theta_L + 2(1-\lambda) r(\theta - \theta_L)\right]}\\ &= \frac{\pi_0 - \bar{\pi} + \beta a \theta_L\left[\theta_L + (1-r)\chi(\theta_H - \theta_L)\right]}{\beta a \theta_L\left[(\beta/\gamma)\theta_L + 2(1-r)\chi(\theta_H - \theta_L)\right]}\end{aligned}$$

be decreasing in r, where $\chi \equiv q/(1-q)$. This occurs when

$$\begin{aligned} \begin{vmatrix} \beta a \theta_L & \pi_0 - \bar{\pi} + \beta a \theta_L^2 \\ 2 & (\beta/\gamma)\theta_L \end{vmatrix} &= 2\left(\bar{\pi} - \pi_0\right) - \left(2 - \frac{\beta}{\gamma}\right)\beta a \theta_L^2 > 0 \iff \\ \left(1 - \frac{\beta}{2\gamma}\right)\theta_L^2 &< \frac{\bar{\pi} - \pi_0}{\beta a},\end{aligned}$$

hence the result under Assumption 1. ■

Proof of Proposition 3

Let us examine the incentive to repress (gross of memory costs):

$$\begin{aligned} \tilde{U}(\pi, r, \mu; \lambda, r) - \tilde{U}(\pi, 0, \mu; \lambda, r) &= a\beta(1-\tau)^2\theta_L\left(\theta(r) - \theta_L\right) - a\beta^2(1-\tau)^2\left(\frac{\theta(r)^2 - \theta_L^2}{2}\right)\\ &= a\beta(1-\tau)^2\left(\theta(r) - \theta_L\right)\left[\theta_L - \beta\left(\frac{\theta(r) + \theta_L}{2}\right)\right]\\ &= a\beta(1-\tau)^2\left(\theta_H - \theta_L\right) r\left[(1-\beta)\theta_L - \beta r\left(\frac{\theta_H - \theta_L}{2}\right)\right] \quad (A.10)\end{aligned}$$

[9] If tax rates were constrained to be nonnegative, we would have instead $\tau^i = \max\left\{\min\left\{T(\pi, \mu^i; \lambda, r), 1\right\}, 0\right\}$; this would make little much difference to the results.

The required equilibrium conditions are therefore that:

$$\beta \underline{r}\left(\frac{\theta_H-\theta_L}{2}\right) \quad < \quad (1-\beta)\theta_L \tag{A.11}$$

$$(1-\bar{\tau})^2\bar{r}\left[(1-\beta)\theta_L-\beta\bar{r}\left(\frac{\theta_H-\theta_L}{2}\right)\right] \quad < \quad (1-\underline{\tau})^2\underline{r}\left[(1-\beta)\theta_L-\beta\underline{r}\left(\frac{\theta_H-\theta_L}{2}\right)\right] \tag{A.12}$$

Since $(1-\bar{\tau})^2<(1-\underline{\tau})^2$, the second condition is satisfied when

$$\begin{aligned}(1-\beta)\theta_L\left(\bar{r}-\underline{r}\right) \quad &< \quad \beta\left(\theta_H-\theta_L\right)\left(\frac{\bar{r}^2-\underline{r}^2}{2}\right) \iff \\ (1-\beta)\theta_L \quad &< \quad \beta\Delta\theta\left(\frac{\bar{r}+\underline{r}}{2}\right).\end{aligned}$$

Thus, the two requirements jointly take the following form:

$$\beta\left(\frac{\Delta\theta}{2}\right)\underline{r}<(1-\beta)\theta_L<\beta\left(\frac{\Delta\theta}{2}\right)\left(\bar{r}+\underline{r}\right). \tag{A.13}$$

Note: may want to how to find a condition that implies it and is independent of $\left(\underline{r},\bar{r}\right)$. An "obvious" one would be $(\beta/2)(\Delta\theta)<(1-\beta)\theta_L<\beta(\Delta\theta)q$, but this requires $q>1/2$, which we may not want to impose. Alternatively, pick β in the interval:

$$1+\left(\frac{\Delta\theta}{2\theta_L}\right)\underline{r}<\frac{1}{\beta}<1+\left(\frac{\Delta\theta}{2\theta_L}\right)\left(1+\underline{r}\right), \tag{A.14}$$

Finally, when $\bar{\lambda}<1$ we also need to check also that no agent want to rehearse the bad news:

$$\tilde{U}_L(\pi_0,\bar{\tau},\bar{r};\bar{\lambda},\bar{r})-\tilde{U}_L(\pi_0,\bar{\tau},0;\bar{\lambda},\bar{r}) \quad > \quad -m'(1-\bar{\lambda}), \tag{A.15}$$

$$\tilde{U}_L(\pi_0,\underline{\tau},\underline{r};\underline{\lambda},\underline{r})-\tilde{U}_L(\pi_0,\underline{\tau},0;\underline{\lambda},\underline{r}) \quad > \quad -m'(1-.\underline{\lambda}). \tag{A.16}$$

Given (24), the second condition is ensured by the fact that the left-hand side is strictly positive. From (A.10), the first condition is satisfied provided that

$$\beta\bar{r}\left(\frac{\Delta\theta}{2}\right)<(1-\beta)\theta_L, \tag{A.17}$$

or of course if m' is large enough. Note that (A.17) tightens the double inequality in Assumption (4) still further. With respect to m', we probably want to have (or at least not exclude), $m'<m$ (except perhaps when $\bar{\lambda}$ is close to 1).\It only remains to show that Assumptions 1–4 define a nonempty region of the parameter space. [To do]. ■

REFERENCES

Akerlof, G. and W. Dickens, "The Economic Consequences of Cognitive Dissonance," *American Economic Review,* LXXII(1982), 307–319.

Alesina. A. and La Ferrara, E. (2001) "Redistribution in the Land of Opportunities," NBER Working Paper...

Alesina, A., Glaeser, E., and Sacerdote, B.(2002) "Why Doesn't the US Have a European-Type Welfare State?" *Brookings Papers on Economic Activity...*

Bénabou, R. (2000) "Unequal Societies: Income Distribution and the Social Contract," *American Economic Review,* 90, 96–129.

————and Ok, E. (2001) "Social Mobility and the Demand for Redistribution: the POUM Hypothesis," *Quarterly Journal of Economics,* 116, 447–487.

———— and Tirole, J. (2002) "Self–Confidence and Personal Motivation," *Quarterly Journal of Economics,* August, forthcoming.

———— and Tirole, J.(2000) "Willpower and Personal Rules," Princeton University mimeo, June.

Bisin, A. and Verdier, T.

Brocas, I. and Carrillo, J. (2000) "Entrepreneurial Boldness and Excessive Investment," CEPR Discussion Paper, 2213.

Carrillo, J., and T. Mariotti (200) "Strategic Ignorance as a Self-Disciplining Device," *Review of Economic Studies,* 66, 529–544.

Hirschman, A. O. (1973)"The Changing Tolerance for Income Inequality in the Course of Economic Development (with a Mathematical Appendix by Michael Rothschild)" *Quarterly Journal of Economics*, LXXXVII , 544–566.

Hassler J., Rodriguez–Mora, J., Storesletten, K., and Zillibotti, F.(2002) "The Survival of the Welfare State", *American Economic Review...*

Hochschild, J. (1991)

———— (1996)

Keely, L. (2002) "Mobility and Fertility," University of Wisconsin-Madison mimeo.

Ladd, E. and Bowman, K. "Attitudes Towards Economic Inequality," American Enterprise Institute Studies on Understanding Economic Inequality.

Lamont, M. (2002)

Lane, R. (1959) " Fear of Equality," *American Political Science Review,* 59(1), 35-51.

Lane, R.

Lerner, M. (1982) *The Belief in a Just World: A Fundamental Delusion. New York, NY: Plenum Press*

Piketty, T. (1995a) “Social Mobility and Redistributive Politics,” *Quarterly Journal of Economics*, CX, 551–583.

———— (1998) “Self-Fulfilling Beliefs About Social Status,” *Journal of Public Economics*,70, 115-132.

———— (1995b). “Redistributive Responses to Distributive Trends,” MIT mimeo.

Putterman, L, J. Roemer, and J. Sylvestre, “Does Egalitarianism Have a Future?” *Journal of Economic Literature*, XXXVI (1996), 861-902.

Roemer, J., “Why the Poor Do Not Expropriate the Rich in Democracies: An Old Argument in New Garb,” *Journal of Public Economics*, LXX (1998), 399–426.

Rotemberg, J. “Perceptions of Equity and the Distribution of Income,” *Journal of Labor Economics...*

Saint Paul (2001) “The Dynamics of Exclusion and Fiscal Conservatism,” *Review of Economic Dynamics...*

Shapiro, R.

Suhrcke, M. (2001) “Preferences for Inequality: East Versus West,” Innocenti Working Paper N0. 89, UNICEF, Florence.

Vindigini (2002).