The Dynamics of Environmental Politics and Values*

Timothy Besley
LSE

Torsten Persson
IIES, Stockholm University

May 6, 2019

Abstract

This paper develops a framework to study environmentalism as a cultural phenomenon, namely as reflecting a process of social identification. The model.is used to explain how the shares of environmentalists and materialists in society can coevolve with taxes on emissions to protect society against damages by environmental degradation. These policies are determined by electoral competition. However, even though politicians internalize the welfare of those currently alive and pick Utilitarian optimal policies, the dynamic equilibrium paths of policies and evolving values may not converge to the steady state with the highest level of long-run welfare.

*This paper is based on Besley's 2017 FBBVA Lecture. The authors are grateful for input from the audiences when the lecture was given at the ASSA meeting in Chicago and the FBBVA headquarters in Madrid, and for perceptive comments by Bård Harstad. Financial support from the ERC and the Swedish Research Council is gratefully acknowledged.

1 Introduction

Understanding the process whereby policy is determined is an important reason to study political economics. The models and approaches that have been developed in this field can help identify a range of constraints on policy-making that explain the failures to adopt better policy outcomes. Depending on the application, these constraints can emanate from the policy preferences of various actors, the extent of optimizing behavior by these actors, the technologies used in private and public production, the information and commitment abilities available to policymakers, and the institutional arrangements that allocate the use of political power.

The existing political-economics literature has mainly put the spotlight on the institutional framework. To the extent it considers dynamic factors, these typically relate to changes in wealth, while policy preferences are treated as fixed. Given these preferences, policies shape payoffs over time by changing incentives rather than from systematically changing preferences or values.

Arguably, this is a restrictive model for many policy issues. However, endogenizing societal values as manifested in preferences remains a controversial topic and – as a result – economists rarely try to unpack the factors that drive preferences. In many cases, policies have to respond to social values and can also help shape those values. One such case is the topic of this paper: the determination of environmentalist values.

To study this issue requires breaking with the strong economics tradition of focusing solely on materialistic preferences. We thus conceive of environmentalism as a fundamental value with consequences for consumption behavior – valuing a life style which limits pollution. A major reason to study such values is today's concern about man-made climate change. Without substantial reductions in carbon emissions, we may eventually raise the risks of severe disruptions of our lifestyles and even the risks of human extinction. Hence, many people who regard themselves as environmentalists try to reduce their carbon footprint. Moreover, they frequently do so in very visible ways that signal their values to others, such as driving an electric vehicle or riding a bicycle to work.

Economists tend to argue that the best remedy for combatting climate change, or other pollution damages, is to change policy, by imposing a Pigouvian tax on fossil fuels or by fine-tuning a system of tradable quotas for carbon emission. But such recommendations may not reach very far, unless we recognize that policy is the product of a political process – it is an equilibrium

outcome, rather than a primitive. The real obstacle to imposing desirable policies may thus lie with government, even though what government does could reflect the views of their citizens, especially in democratic societies.

Climate activists tend to take another approach to finding a remedy for climate change, namely to influence people's values. Changing values may operate directly by changing behavior. But they may also operate indirectly by changing the outcomes of the political process. More interesting still is the possibility that modifying today's values may change policies, which feed back to tomorrow's values, thus creating a self-enforcing process of change. However, to explore these channels one needs to treat values as endogenous.

Our paper develops a basic framework to analyze the interacting dynamics of values, politics, and environmental policy. A key feature of the model is to identify theoretically a range of complementarities between value adaptation and policy choice. We argue that such dynamics can have important implications for the long-run patterns of social change, where environmentalist values either grow or diminish endogenously over time. Our main focus is on macro-trends in preferences, when the key dynamics arise via a political externality due to the working of democratic society. This focus is rooted in facts – we show that societies display very significant differences in environmental attitudes that cannot be accounted for by differences in individual characteristics.

Our model is very simple and has two kinds of citizens: materialists and environmentalists. Policy is determined by electoral competition between two office-seeking parties that court citizens who are willing to swing their votes. As a result, they set the environmental policy to cultivate the interests of the average swing voter – this results in a Pigouvian-like policy outcome. Relying on a very simple evolutionary process, we show how these policy choices can drive a society towards either environmentalism or materialism.

We also study the long-run welfare consequences of such changes in societal values. If environmental damages are large enough, a society's welfare is the highest when its population consists of only environmentalists, as this leads to the eradication of pollution. We argue that failure of a democratic political system to achieve this outcome reflects the system's inability to commit. This inability gives a potential welfare-enhancing role for independent institutions insulated from politics, assuming that society can stand by its commitment to these institutions.

The remainder of the paper is organized as follows. The next section discusses some related literature. Section 3 justifies our focus via some facts

about individual-level and country-level environmental values from the World Values Survey. Section 4 lays out the economic and political sides of our theoretical framework for a given share of environmentalists in society. Section 5 develops our dynamic model of changing values. Section 6 studies the welfare implications of our modeling. Section 7 mentions a few possible extensions of the analysis. Section 8 concludes.

2 Related Literature

The analysis in this paper is related to different bodies of research. A key dimension of the analysis is indeed to cross-fertilize ideas from different parts of social science: economics, politics, sociology, psychology, and anthropology.

Policy responsiveness in static models By now, a large and established theoretical and empirical literature shows how elections or pressure groups may shape policy within a given set of political institutions. Most of the models are static, studying how policies are chosen to affect the current generation of citizens with no implications for future policy. The main issue in such models is who gets what out of the political process. For instance, the classic model of Downs (1957) predicts that parties motivated mainly by winning an election adopt the preferred position of the median voter, if such a position exists.

Difficulties with equilibrium existence – except in very stylized policy-making environments – led to the development of models with shocks to voters, such that the winning probability varies smoothly with policy positions. Such probabilistic voting models have been used extensively (Coughlin 1992, Lindbeck and Weibull 1987, or Persson and Tabellini 2000). Citizen-candidate models (Besley and Coate 1997, or Osborne and Slivinski 1996) similarly develop a static framework where policy is chosen by elected officials, but where these now represent different groups in society. Although the exact mechanism varies, a key feature of all these models is that policy responds citizens' policy preferences. Indeed, some would say that such responsiveness is the *sine qua non* of democracy.

The standard normative benchmark used in models of environmental economics is a Pigouvian tax which is set equal to the pollution externality associated with the marginal damage done to the environment. Following in the same tradition, the standard economic approach to environmental-

ism (reviewed by Oates and Portney 2003) supposes that underlying values and preferences are fixed. Instead, it examines a static setting, where interest groups lobby policy makers to move policy in their preferred direction. These policy makers have mixed motives over social welfare and money transfers, which are presumably related to the citizens' underlying preferences as expressed through elections.

Dynamic political-economics models Turning to dynamics, a number of strategic models where current policy choices can influence future elections or policy outcomes. The first main application of dynamic political models was to explore dynamic inter-linkages from debts and deficits. Thus, Persson and Svensson (1989) argue that debt policy can stray from the efficient path in political equilibrium, while Tabellini and Alesina (1990) show how political instability can give incumbent officeholders incentives to borrow from the future. In both these papers, current incumbents strategically alter future debt levels to manipulate the choices of future policy-makers who may not share their own policy preferences.

Aghion and Bolton (1990) and Milesi-Ferretti and Spolaore (1994) develop models where policy is distorted because current policy choices affect which political party will win in the future. The use of policy in such models is deliberate and strategic and politicians understand that today's policies have future political implications. A similar effect has been used to explain other policy puzzles. For instance, Acemoglu and Robinson (2006) ask why some governments fail to invest in developing the economy and refer to a "political replacement effect," whereby today's policymaker fears that certain policies reduce her chances to survive in office.

Acemoglu (2003) and Besley and Coate (1998) point to a key feature of dynamic political models, namely that policies can be rendered inefficient by an inability to commit by political decision-makers. For example, in a two-group setting, a policy which could benefit both groups may not be implemented because one group cannot commit to compensate another in the event of a transition in political power. A lack of commitment power in the political process gives way to a potential role for non-elected independent institutions to act as a source of credible commitment. The classic example is the use of an independent central bank to make monetary policy when politicians are tempted to inflate the economy as in Rogoff (1985), while Acemoglu and Robinson (2000) point to the introduction of the franchise

when incumbent elites are tempted to renege on promises of redistribution to the masses.

Yet another strand of dynamic modelling is developed in the literature on state capacity. Besley and Persson (2009, 2011) study models where the main dynamic force comes from the state investing in its own functions – the powers to tax, to adjudicate, and to deliver public goods. This analysis shows that finding institutional ways for more cohesive politics can serve as to spur such investments.

None of these dynamic models consider changing values and preferences as an explicit mechanism. To the extent policy preferences change over time, this reflects changing economic interests of citizens with fixed preferences.

Values, preferences, and identities We will think of environmentalism as a particular form of pro-social preferences – i.e., as a commitment to a specific cause. This means that our model of environmentalists versus materialists relates to interconnected literatures on intrinsic motivations. It is well known that private intrinsic motivations to do good can underpin pro-social actions, e.g., in charitable giving (Andreoni 2006). We model environmentalists as people who take a certain action: they choose not to consume polluting goods. Even though an individual environmentalist cannot materially affect the level of pollution with her own actions, she contributes to the environmental cause. Other ways to think about the same phenomenon would run via mission-driven preferences (Besley and Ghatak 2005), or via adoption of a particular social identity (Akerlof and Kranton 2005, 2010). Along the latter way, the identity of being an environmentalist includes not to consume goods that cause pollution.

We will adopt a specific micro-foundation for this behavior based on Benabou and Tirole (2006). Thus we suppose that the motive to become an environmentalist is to get social respect as a virtuous person. However, in our version of their social-signalling model, consumption is only imperfectly observed. This means that the value of social signals depends on the identity shares in the population. With a high fraction of environmentalists, it is more likely that someone not observed to consume polluting goods is in fact an environmentalist. By contrast, with a low share of environmentalists, there is little signalling value in abstaining from consuming polluting goods.

Cultural dynamics Generally speaking, the paper is part of a wider agenda aimed at studying the coevolution of values or cultures with other strategically designed outcomes. There is little doubt that drivers of preferences and values not only reflect inherited genetic endowments but are also shaped by cultural fitness. Cultural "memes" can be propagated by social influences transmitted by families, teachers, peer groups and other social networks. Such social factors surely act on us throughout the life cycle, but they may be particularly important during the formative years of childhood and adolescence. Social influence at that critical stage of life may thus leave a permanent mark on choices and behavior in adulthood.

Taking explicit cultural dynamics seriously is a recent development in economics where culture has generally been thought of as an "error term". But economists increasingly appreciate that certain cross-sectional and time-series observations of cultural traits cry out for explanation, using economics tools and methods of empirical investigation.

Resistance to these ideas among economists create barriers to dialogue across disciplines. Thus, the idea that preferences are fluid and socially determined is readily accepted among sociologists. In a classic account, Bales and Parsons (1955) put it as follows:

> "If .. the essentials of human personality were determined biologically, independent of social systems, there would be no need for families ... It is because the human is not "born" but must be "made" through the socialization process that ... families are necessary. They are "factories" which produce human personalities.... We therefore suggest that the basic and irreducible functions of the family are two: first the primary socialization of children so that they can truly become members of the society ... ; second, the stabilization of the adult personalities of the society." (pages 16-17)

Among economists, Sam Bowles is a pioneer for this view (Bowles 1998):

> "the argument that economic institutions influence motivations and values is plausible, and the amount of evidence consistent with the hypothesis is impressive. Many ethnographic and historical studies, for example, recount the impact of modern economic institutions on traditional or indigenous cultures. The rapid rise of feminist values, the reduction in family size, and

> the transformation of sexual practices coincident with the extension of women's labor force participation likewise suggest that changes in economic organization may foster dramatic changes in value orientations." (page 76)

The ideas in this paper are also heavily influenced by the formal models of cultural change developed by evolutionary anthropologists (Boyd and Richerson 1985, Cavalli-Sforza and Feldman 1981). They borrow from the formal structures of population biology to model behavioral change as social learning that propagates behavior across populations. This approach has been influential in exploring the basis of unselfish behavior in kin groups or broader social groups. An important idea in this literature is the notion of cultural parents who influence the behavior of their offspring. Cultural parents are not confined to biological or foster parents, and can include a wide range of peers in education, social life, and education.

A canonical example in this research is the public-goods game, where rational self-interested individuals do not contribute because of their incentive to free-ride on the contributions of others. But this can be altered by three evolutionary mechanisms: mutations, genetic drift (relevant only in finite populations), and natural selection. Boyd and Richerson (1982) consider "conformist transmission" where individuals imitate the more common behavioral types among their cultural parents, which raises the frequency of these types in the population.

A growing literature in economics considers related ideas – see Bisin and Verdier (2011) for a useful review[1]. Bisin and Verdier (2001) develop a model where parents strategically socialize their children, by weighing future payoffs of children against the "social distance" between parents and children. Bisin and Verdier (2000) apply this approach to study the dynamics of religion and ethnicity. Kuran and Sandholm (2008) also develop a model of cultural integration. Bezin (2015) proposes a model of cultural evolution for environmental preferences based on private contributions to environmental protection.

The evolution of preferences we use in this paper builds on the indirect evolutionary approach introduced in Güth and Yaari (1992) and Güth (1995), who propose that preferences respond to payoffs in repeated games. In such models, whether or not preferences are observable is a key issue. In our

[1] See also Saez-Marti and Zillibotti (2008) for an overview of the issues.

setting, this is not important because the key externality runs through an electoral process, where individuals (stochastically) vote in a sincere way. Ostrom (2000) emphasizes the indirect evolution approach in a context of collective action. By changing preferences, societies can become more or less cooperative and hence more or less able to solve collective-action problems. To date, however, most applications of these ideas have focused on small-scale cooperation.

Institutions and preference dynamics Analyzing how policy influences preferences and values has not been explored much at all. Besley (2017) develops a model where the choice of redistributive policies affects the aspirations of citizens which in turn affects future policy. His model allows preferences and redistribution to coevolve along the equilibrium path. Besley and Persson (2019a) build a multi-dimensional dynamic model to explore identity politics in an attempt to explain the rise of nationalist "identity politics". Besley (2019) explores how compliance with taxation depends on preference types in the population with a dynamically evolving level of taxation. Besley and Persson (2019b) ask how the design of democratic institutions interacts with democratic values over time. The general coevolution of institutions and culture is explored in Bowles, Choi, and Astrid (2003) and Bisin and Verdier (2017).

3 Environmentalist values

As already mentioned, those concerned about human impacts on the environment often suggest that changing values provide a route to more sustainable behavior and policy. Environmental values are most commonly expressed through spending patterns or recycling, but also through political activism. One outstanding issue is whether the underlying values are based on self-interest, humanistic altruism, or biospheric altruism. Dietz et al (2005) provide a systematic review of these issues across the social sciences. The authors discuss how environmental values can be thought as a kind of "post-materialist" ethic that can be associated with altruism. However, they lament that

"little can be said about the causes of value change and of the overall effects of value change on changes in behavior." (page 335).

Lorenzoni and Pidgeon (2006) discuss a range of polling data and note

that environmental attitudes vary across populations, particularly between Americans and Europeans. In his overview of the growing social movements that put environmental values at the heart of campaigns to change policy, McAdam (2017) discusses why such values have spawned so little activism in the US.[2]

One of the key ideas in the theoretical model developed in the next section is that people have heterogeneous values regarding the need to protect the environment and that these attitudes shape their policy preferences. To shed some light on these attitudes and their differences across individuals and countries, we turn to the World Values Survey (WVS).

Using the WWS data We use two questions from this survey. The first question appears in WVS waves 3, 4, 5 and 6 and is answered by about 250,000 people. It asks each respondent whether they would prioritize the environment over economic growth, or vice versa. We code the answer in a binary fashion, and set an indicator equal to one if s/he regards protecting the environment as the priority. We think of such individuals as self-identifying as environmentalists. This applies to 54 percent of the full, worldwide sample, although – as we will soon see – this identification varies systematically across countries and across individual characteristics.

The second question is posed in WVS waves 2, 3, 4 and 5 and is answered by about 190,000 people. It concerns policy by asking the respondent about an "increase in taxes if used to prevent environmental pollution." The four alternative answers are: "strongly agree", "agree", "disagree", or "strongly disagree". We code the response as favorable to environmental taxes if the respondent strongly agrees or agrees. This is true for about 44 percent of the sample. Once again, we will find stark individual and cross-country variation

We expect people to express different environmental attitudes depending on the period and circumstances in which they were socialized, and that this socialization predominantly occurs at the earlier, more formative, stages in life.

Variation across individuals and countries To explore this expectation, we first construct a variable for birth cohort, for each ten-year period

[2] Aghion et. al. (2019) explore the role of environmental attitudes in encouraging incentives to innovate and show that such values interact with competition in promoting innovation.

since the 1910s. As education is likely to influence an individual's attitudes to the environment, we also examine this variation using the WVS classification into three levels of education. Figure 1 includes two bar charts that show deviations from the country-specific mean of the answers to our two attitudinal questions. This isolates the idiosyncratic variation by cohort and education group. The left panel shows a clear variation: environmentalism, as well as the willingness to raise environmental taxes, are stronger among more recent cohorts. This suggests a shift in values across generations towards increasing environmentalism. The right panel shows a larger concern for the environment and a greater willingness to put up taxes among more highly educated groups.

Figure 2 displays four cross-country histograms for the share of environmentalists and the share of people willing to raise environmental taxes, defined by our binary classifications. To maximize the number of countries in the data, we average this across all WVS waves. The top two panels show the raw data, while the bottom two panels condition on a range of individual characteristics.[3] Responses are strikingly different across countries, whether we condition on individual characteristics or not. The share of environmentalists e.g., varies between 20 and 80 percent in the raw data and remains highly variable as we condition on individual traits. It is plausible to attribute these macro differences to different cultural values.

Figure 3 shows how our indicators for values and policy preferences covary across countries. It plots the average willingness to raise environmental taxes against the importance of protecting the environment, with the raw data in the left panel and the residuals used in Figure 2 in the right panel. Both graphs display a clear positive correlation, showing that values and policy preferences indeed go hand in hand. The correlation is strong, especially if we ignore the outliers on (both sides of) the horizontal axis. These graphs also suggest a country-specific component of social values, or culture, in the environmental domain.

Taken together, Figures 1-3 convey two salient patterns in the data. Environmental attitudes clearly vary across cohorts, as well as categories of education. They also display substantial macro, country-level, differences which cannot be explained by individual characteristics. These differences

[3] We estimate a linear probability model at the individual level with either environmental dummy on the left-hand side and right-hand side variables for gender, ten dummies for income groups, three for education groups, three age bands and wave dummies. To construct the bottom-row histograms, we average the residuals at the country level.

across generations and societies give some underpinnings to our modeling, which indeed implies society-specific environmental values that change systematically across generations.

4 Static Economics and Politics

There are three main elements in our modeling approach. The first is economics, where we use a simple model with two types of citizens/consumers, who face an environmental policy. The second is politics, where we use a probabilistic-voting model of two parties that try to win elections with swing voters and loyal voters among the citizens. These elements are described in this section. As in Besley and Persson (2019a), the more novel part of the model is its third element, namely the evolutionary component whereby values (citizen/consumer types) evolve over time in response to (expected) policy. That element is described in the next section.

Basics Time is infinite and is labelled by s. For most purposes, we will think of s as labeling a sequence of generations. Thus only a single generation is alive at each date. The population includes two types of citizens denoted by $\tau \in \{m, e\}$ where m stands for materialists and e for environmentalists. Let μ_s be the proportion of type m individuals in the population at date s.

4.1 Economics

Everybody in society has the same level of income y. Citizens choose a life style which determines how much they pollute the environment with their consumption. We associate pollution – think about carbon emissions – with a specific component of consumption, a good denoted by c which can be taxed at rate t. There is also another type of non-polluting (ecological) consumption, n, with price p. We can think of $p > 1$ as the normal case – i.e., consuming non-polluting goods is generally more expensive. Consumers are of two types: materialists and environmentalists.

Materialists Materialists have preferences

$$u^m = \log(Ac) + n - \lambda C,$$

which are linear in the numeraire, n, and where the final term is the disutility from environmental damage which we assume depends on average per-capita consumption of the polluting good denoted by C. It is easier to pursue the analysis in terms of $\alpha = \log(A)$,a parameter that additively shifts materialist preferences and therefore plays a role in the utility comparisons between types which we conduct below.

The budget constraint for consumption is $y + r = c(1+t) + pn$, where r is a lump-sum government transfer. Optimal polluting consumption is given by

$$\widehat{c}(t,p) = \arg\max_{c} \left\{ \alpha + \log(c) + \frac{y+r}{p} - \frac{(1+t)c}{p} \right\} = \frac{p}{(1+t)}.$$

As each consumer is small, she cannot affect C by her own actions and hence ignores the effect of consumption on overall pollution. We assume that $0 < \widehat{c}(t,p) < \frac{y}{p}$ and let

$$v(t,p) = \alpha + \log(\widehat{c}(t,p)) - \frac{(1+t)\widehat{c}(t,p)}{p} = \alpha - 1 + \log\left(\frac{p}{(1+t)}\right) \quad (1)$$

be the indirect utility function from good c.

Environmentalists Environmentalists get utility from social signalling as well as from consuming the numeraire. Specifically, they have preferences

$$u^e = n - \lambda C + V(\mu)$$

and, as with materialists, face a budget budget constraint $y = c(1+t) + pn$. Since environmentalists get no utility from the polluting good, they will optimally set $c = 0$.

Apart from non-polluting consumption, the utility function has two additional components. As materialists, environmentalists suffer from environmental damages, according to $-\lambda C$. The second component is a measure of private virtue, a "warm-glow" from being an environmentalist. This utility component reflects being thought of as a good person and – for the sake of concreteness – we model this in the social-signalling framework of Benabou and Tirole (2006). We assume that types are never directly observed. Specifically, if a consumer sets $c = 0$, this can be observed. However, $c > 0$ is not observed with probability $\rho \in (0,1)$.

According to Bayes rule, we can write the probability that an individual who is observed setting $c = 0$, is an environmentalist:

$$\varphi(\mu) = \frac{\mu}{(1-\mu)\rho + \mu}.$$

Note that $\varphi(\mu)$ is increasing in μ with $\varphi(1) = 1$ and $\varphi(0) = 0$.

We suppose that the virtue utility from being an environmentalist is $V(\mu) = \chi\varphi(\mu)$ where $\chi > 0$ is the reputational gain from being thought of as an environmentalist rather than a materialist. Assuming that χ is positive is a simple way to microfound why it is attractive to be an environmentalist: although you forgo some private consumption, you gain social respect. This corresponds to what is often called "virtue signalling". Our specific micro-foundation generates a *positive* link between utility from virtue and μ – in a society with many (few) environmentalists, it is more (less) likely that someone not observed choosing $c > 0$ is indeed an environmentalist.[4] This link will ensure that virtue signalling can have social effects by generating benefits to being an environmentalist which can promote cultural change.

Policy preferences To close the model, we assume that tax revenue is rebated back to both groups of consumers on a uniform basis through per-capita (lump-sum) grant:

$$r = Ct.$$

We now substitute from the consumer and government budget constraints, use the equilibrium condition that $C = (1-\mu)\widehat{c}(t,p)$, and set p to 1 so that we can eliminate it as an argument from the functions above.[5] This allows us to write the policy preferences of the two types

$$u^{\tau}(t,\mu) = \begin{cases} \chi\varphi(\mu) - (\lambda - t)(1-\mu)\widehat{c}(t) + y & \text{if } \tau = e \\ v(t) - (\lambda - t)(1-\mu)\widehat{c}(t) + y & \text{if } \tau = m. \end{cases} \tag{2}$$

As $v(t)$ is decreasing in t, the tax rate preferred by environmentalists is strictly higher than that preferred by materialists.

[4] This is one of several possible micro-foundations for warm-glow utility to be increasing in μ.

[5] Setting $p = 1$ is not totally innocuous as it affects the comparison of utilities between the two types given that environmentalists only consume the numeraire.

4.2 Politics

We now specify the policy process, in a given period s, for fixed preferences (types) in the population. As in Besley and Persson (2019a), we can think of the model as portraying two cleavages: party politics and identity politics. We assume that party preferences are partly based on a fixed dimension, such as class or religion, which creates loyalty to a particular party for some groups of voters. (Such preferences have extensively documented in survey data.) Identity politics is a dimension, not perfectly correlated with party – in this instance, voter type τ. Divisions in this dimension affects policy preferences – in this instance, preferences over environmental taxes t – which spill over to preferences over party platforms. In the next section, we will study how non-partisan identity to evolve dynamically.

Parties Consider a model of two-party competition with probabilistic voting. We label the parties A and B and assume that they are solely motivated by winning elections.[6] Each of the two parties chooses a party platform for its proposed environmental tax rate: $\left\{t^A, t^B\right\}$.

We pick this particular formulation for pure convenience. It will also clearly illustrate how our framework departs from standard models by allowing the population types to evolve over time. But the effects of changing types would extend to any kind of model where a higher population share of a certain type moves policy towards the one preferred by that type. In the probabilistic-voting approach, this happens smoothly (see, e.g., Lindbeck and Weibull 1987, or Persson and Tabellini 2000).

Voters There are two groups of voters. Swing voters cast their ballots based on proposed policy platforms. Loyal voters always cast their ballots for one of the parties. This distinction follows a long-standing, political-science tradition based on the Michigan voting surveys. To simplify the algebra, we assume that half of the voters from each group are swing voters.[7]

[6] The fundamental insights would also hold with policy-motivated parties as long as there is some incentive for parties wish to court non-environmentalist swing voters as part of their electoral strategy.

[7] It would be easy to have $\gamma\mu$ type-e swing voters and $(1-\gamma)(1-\mu)$ type-m swing voters where the parameter γ measured the relative tendency for environmentalist voters to be swing voters and a fraction $\gamma+\mu-2\gamma\mu$ of loyal voters are attached 50-50 to the two parties.

Following the probabilistic-voting approach, the party choice by swing voters is also subject to idiosyncratic (voter specific) and aggregate (affecting all voters) shocks. A swing voter of type τ supports party A if

$$u^{\tau}(t^A, \mu) + \varepsilon + \chi \geq u^{\tau}(t^B, \mu),$$

where ε is the idiosyncratic shock and χ the aggregate shock. Both shocks are assumed to be uniformly distributed ε on $[-1/E, 1/E]$ and χ on $[-1/X, 1/X]$. This simple formulation – together with our specific assumptions about individual utilities – gives a closed-form solution for policy.

Given our assumptions and integrating over ε, the proportion of type τ swing voters who vote for party A is

$$\frac{1}{2} + E\left[u^{\tau}(t^A, \mu) - u^{\tau}(t^B, \mu) + \chi\right]. \tag{3}$$

We assume an interior solution – i.e. that (3) lies between zero and one.

Winning probabilities Party A wins the election if it gets more than half of the votes. This will happen if

$$\chi + \Omega\left(t^A, t^B, \mu\right) \geq 0, \tag{4}$$

where

$$\Omega\left(t^A, t^B, \mu\right) = \begin{array}{c} \mu\left[u^e(t^A, \mu) - u^e(t^B, \mu)\right] \\ +(1-\mu)\left[u^m(t^A, \mu) - u^m(t^B, \mu)\right]. \end{array}$$

The first term in (4) just depends on whether the realized aggregate shock χ favors party A, while the second depends on whether the policies on offer allow the party to court swing voters.

Integrating across χ, gives us the probability that party A wins the election as:

$$q^A = \frac{1}{2} + X\Omega\left(t^A, t^B, \mu\right). \tag{5}$$

assuming an interior solution.[8] Party B wins with the complementary probability $q^B = 1 - q^A = \frac{1}{2} - X\Omega\left(t^A, t^B, \mu\right)$. It follows that the probability of winning for each party is given by the same function of its own tax rate. Given the expression for $\Omega\left(t^A, t^B, \mu\right)$, the parties are thus effectively maximizing the same Utilitarian social-welfare function.

[8] This will always be the case if X is small enough – i.e., there is a wide enough support for common shock χ.

Equilibrium tax rates To study equilibrium policy choices, we look for a Nash equilibrium where each party optimizes its policy platform given the decision of the other. Given the comments above, the political equilibrium thus maximizes a Utilitarian objective similar to what would emerge from a standard Pigouvian model. Specifically, we have:

Proposition 1 *Both parties pick the same tax rate to maximize their winning probability*

$$t^A = t^B = \widehat{t}\,(\mu) = \frac{\mu + \lambda}{1 - \mu}. \tag{6}$$

Proof. To prove this, note that and $p = 1$ – (2) implies

$$t = \arg\max \left\{ \frac{(t - \lambda)}{1 + t} + (1 - \mu)\, v\,(t) \right\}$$

So the first-order condition is

$$-\frac{(t - \lambda)}{(1 + t)^2} + \frac{\mu}{1 + t} = 0$$

Solving this expression yields the result. ■

Observations We end this section with three observations about the solution in Proposition 1.

First, the proposition gives the lowest tax rate as $t\,(0) = \lambda$. This is the conventional Pigouvian tax that exactly corrects for the environmental externality in a population where everybody causes pollution. For positive shares of environmentalists, the tax rate is higher. Note, however, that the political equilibrium produces a Utilitarian optimal policy for any given μ. Hence, there is no political failure according to conventional economic thinking. The political equilibrium tax is thus a useful and important benchmark for the analysis to follow. There, we will show that with endogenous values the traditional Pigouvian optimum is not the appropriate optimum, even if we stick to utility-based criteria.

Second, as $\mu \to 1$, we have $t\,(\mu) \to \infty$. Thus the tax rate gets so high that any remaining materialist is completely discouraged from consuming the polluting good. In effect, the tax is prohibitive and therefore equivalent to a ban on consumption.

Third, define

$$u^{\tau}(\mu) = u^{\tau}(\widehat{t}\,(\mu)\,, \mu), \tag{7}$$

as the equilibrium utility of type τ when the population has a fraction μ of environmentalists. It is clear that $u^{e}(\mu)$ is increasing in μ, but $u^{m}(\mu)$ is decreasing in μ. This is because a higher fraction of environmentalists makes politicians put more weight on their preferences relative to those of materialists.

5 Dynamics of Environmentalism

The dynamics of preferences is the least standard element of our analysis. This section specifies the evolutionary model we rely on, analyzes the resulting dynamics and describes the model's steady state(s). Given the binary types and our formulation of the evolution process, this analysis turns out to be simple.

The evolution of values We use a general class of value dynamics, based on "cultural fitness" – i.e., on the comprehensive payoff advantage one type enjoys relative to another given the state of the "culture" at a given date. This embodies the idea that a payoff advantage drives the dynamics of values, which is quite a weak assumption. However, this only one possibility out of several. For example fitness could be purely material – e.g., based on real income levels. Or fitness could be purely social – e.g., based on dominant behavior in a relevant group.

As part of the socialization process, we assume a capacity for intrapersonal comparisons of utility between materialists and environmentalists as, for example, when parents are assessing the hypothetical psychological well-being of socializing their children as alternative types.[9]

To model the evolutionary process, we follow Sandholm (2010), who suggests an evolutionary dynamic where individuals may change their types sporadically (inertia) and base their switch on current behavior and opportunities (myopia). This approach is underpinned by a revision protocol which, formally, is a continuous function: $\varsigma_{i,j}\left(u^{i}(\mu), u^{j}(\mu), \mu\right) \in [e, m]$ that specifies a conditional switch rate from type i to j given the payoffs and proportion of

[9] This is true throughout the literature e.g. Bisin and Verdier (2011) makes this assumption.

types in the population. Sandholm (2010) suggests a general class of mean dynamics such that:

$$\mu_{s+1} - \mu_s = (1 - \mu_s)\,\varsigma_{e,m} - \mu_s \varsigma_{m,e}, \tag{8}$$

where

$$\varsigma_{i,j} > 0 \iff u^j(\mu_{s+1}) - u^i(\mu_{s+1}) > 0.$$

This class of dynamics is convenient, since the direction of change depends only on comparing the payoff from being one type with the potential payoff from being the other type. The essential ingredient of this model is that a type in the population that "thrives" psychologically will tend to grow in a way that depends on the payoff difference. With (8), the driver of preference dynamics depends upon the utility difference between environmentalists and materialists

$$\Delta\left(\mu_{s+1}\right) = u^e(\mu_{s+1}) - u^m(\mu_{s+1}).$$

This function is increasing in μ by the observations about (7) at the end of Section 4.

An example of micro-foundations To forge a link with models of cultural dynamics,[10] let us work with a specific micro-foundation of (8). As in Besley (2017), suppose there are successive generations with two generations alive at each date: parents and children. In the current setting, only parents vote (and also make consumption decisions on behalf of their children). To keep the population balanced, every family has two parents and two children. Suppose there is a matching process in which fraction β of mating is assortative such that parents have the same type. The remaining fraction $1 - \beta$ of parents are randomly matched.

Children are socialized by their parents. For simplicity, suppose that two parents of the same type guarantees that their common type is passed along to their children.[11] However, whether a child with mixed and forward-looking parents becomes an environmentalist depends on $\Delta\left(\mu_{s+1}\right)$, i.e., the utility difference for the two types in the next period – when the child is adult. The child's type also depends on a family-specific shock v that has infinite

[10] See Bisin and Verdier (2011) for a review.

[11] This is clearly a strong assumption, adopted here to make the analysis sharper and simpler. One could consider alternatives, such as a fixed "mutation" rate in homogenous groups.

support and distribution function $G(\cdot)$,which is symmetric around a zero mean with density $g(\cdot)$. If the condition for becoming an environmentalist is $\Delta\left(\mu_{s+1}\right) \geq \upsilon$, the probability that an individual with mixed parents becomes an environmentalist is $G\left(\Delta\left(\mu_{s+1}\right)\right)$. Given a continuum of families, this will also be the proportion of environmentalists among those with mixed parents. Note that $G(\cdot)$ increases smoothly in Δ with $G(0) = 1/2$.[12]

In this example, we can write the proportion of the population who are environmentalists at date $s+1$ given that μ_s are environmentalists at s as:

$$\mu_{s+1} - \mu_s = 2\mu_s(1-\mu_s)(1-\beta)\left[G\left(\Delta\left(\mu_{s+1}\right)\right) - \frac{1}{2}\right], \tag{9}$$

which is a special case of (8). To interpret this, note that assortative matching preserves the proportion of environmentalists. However, among the randomly matched, a fraction μ_s^2 are matched with other environmentalists. The fraction of mixed-parent households is therefore $2\mu_s(1-\mu_s)$.[13]

Obviously, the revision protocol in (8) can support a wider set of socialization processes, including those that work via peer groups, teachers, and other cultural parents. But all of them have in common that a psychological-fitness advantage of a type makes it more likely that individuals will belong to this type.

Timing The timing of the dynamic model is as follows

1. There is an initial share of environmentalists in the population represented by μ_s.
2. Parties choose policy and compete for office leading to a tax rate t_s (as described in Section 4.2)
3. Payoffs of citizens are realized (as described in Section 4.1).
4. Citizens match, a new generation is born and children are socialized leading to μ_{s+1}.

[12] Then $\varsigma_{e,m} = \mu(1-\beta)\left[G(\Delta(\mu)) - \frac{1}{2}\right]$ if $G(\Delta(\mu)) < \frac{1}{2}$ and $\varsigma_{m,e} = (1-\mu)(1-\beta)\left[G(\Delta(\mu)) - \frac{1}{2}\right]$ if $G(\Delta(\mu)) > \frac{1}{2}$

[13] Note that the fraction of the population that matches assortatively does not affect the steady state of the model only its speed of convergence as long as $\beta < 1$, i.e. there is some random matching. Parameter β can be thought of as crudely measuring the openness of social structures, as assortative matching will tend to entrench an existing culture while lower β implies more rapid cultural change.

A political complementarity We have seen that the key driver of the dynamics is $\Delta(\mu)$. Using (2), we have

$$\begin{aligned} \Delta\left(\mu_{s+1}\right) &= u^e(\mu_{s+1}) - u^m(\mu_{s+1}) \\ &= \chi\varphi\left(\mu_{s+1}\right) - v\left(\widehat{t}\left(\mu_{s+1}\right)\right). \end{aligned} \tag{10}$$

Note that this expression depends on future equilibrium taxes, which themselves depend on the future state variable μ_{s+1}. Since future taxes will be set in the next period's political equilibrium, it is taken as given by parents and current parties alike. It is straightforward to see that a higher value of μ_{s+1} favors the environmentalists:

$$\Delta_\mu(\mu) = \chi\varphi_\mu(\mu) + \widehat{c}(\widehat{t}(\mu))\widehat{t}_\mu > 0.$$

The first term is positive as the social-signalling mechanism is more effective for a higher value of μ. The second term is positive because a higher μ favors environmentalists in politics and this gives lower indirect utility for materialists via higher taxation (and hence a higher relative price of polluting goods). Thus there is a complementarity between (expected) policy and preference evolution, due to the political externality in majoritarian policy making.

Dynamics and steady states for values Let us make two key observations. First as $\mu \to 0$, $\varphi(\mu) \to 0$ and hence $\Delta(0) < 0$. In words, there is no signalling value from not observing $c > 0$ with no environmentalists in the population, so it is always best to be a materialist. Second as $\mu \to 1$, the $v\left(\widehat{t}(1)\right) \to 0$ and $\Delta(1) > 0$. As polluting goods become prohibitively expensive, the (large) signalling rents from being thought of as an environmentalist are sufficient to give a fitness advantage to environmentalists. Since $\Delta(\mu)$ is continuous and monotonically increasing, there must exist $\hat{\mu}$ defined by $\Delta(\mu) = 0$, such that

$$\chi\varphi(\hat{\mu}) = v\left(\widehat{t}(\hat{\mu})\right). \tag{11}$$

Then, $\Delta(\mu) > 0$ if and only $\mu > \hat{\mu}$. Finally, make the weak assumption that[14]

$$1 - 2\mu(1-\mu)(1-\beta)\, g(\Delta(\mu))\, \Delta_\mu(\mu) > 0 \tag{12}$$

[14] In the second term of the condition, $\mu(1-\mu)$ is maximized at 0.25, while β and g are both smaller than 1. Thus, (12) holds unless Δ_μ is very large.

for all $\mu \in [0,1]$. Since $\Delta(\mu)$ is increasing, (12) implies that any reasonable stability condition makes the candidate interior steady state at $\hat{\mu}$ unstable. The only stable steady states are thus $\mu = 0$ and $\mu = 1$. Using these observations, and letting $\mu_0 \in [0,1]$ denote the initial value of μ, we have:

Proposition 2 *If $\mu_0 > \hat{\mu}$, the polity monotonically approaches the steady state $\mu = 1$. Otherwise, it monotonically approaches the steady state $\mu = 0$.*

Proof. Take a first-order approximation of (9) around μ_s to obtain

$$\mu_{s+1} - \mu_s = 2\mu_s(1-\mu_s)(1-\beta)\left[G(\Delta(\mu_s)) - \frac{1}{2} + g(\Delta(\mu_s))(\mu_{s+1}-\mu_s)\Delta_\mu\right].$$

We can rewrite this expression as

$$\mu_{s+1} - \mu_s = \frac{2\mu_s(1-\mu_s)(1-\beta)}{1 - 2\mu_s(1-\mu_s)(1-\beta)g(\Delta(\mu_s))\Delta_\mu}\left[G(\Delta(\mu_s)) - \frac{1}{2}\right].$$

The denominator on the right-hand side is positive by (12). Since $G(\cdot)$ is increasing with $G(0) = 1/2$, it follows from the term in square brackets that $\mu_{s+1} - \mu_s > 0$ iff $\Delta(\mu_s) > 0$, which requires $\mu_s > \hat{\mu}$. Applying the expression for $\mu_{s+1} - \mu_s$ to $s = 0, 1, 2, ...$ and noting that $\Delta(\mu)$ is monotonically increasing gives the result. ■

To see the logic behind the proposition, note that the environmental tax goes up as the share of environmentalists increases. This, together with our formulation of warm glow μ creates a complementarity between the fraction of environmentalists, μ, and the payoff difference between environmentalists and materialists, $\Delta(\mu)$. The sign of $\Delta(\mu)$, which governs whether μ is increasing or decreasing, switches from negative to positive as μ increases. Which steady state the economy converges to depends on the starting value μ_0 relative to the critical value $\hat{\mu}$, at which $\Delta(\mu)$ switches sign.

Implied policy dynamics The model predicts a changing environmentalist sentiment in politics, governed by changing values. As per the first observation at the end of Section 4.2, this tax is higher than what it would be if it just corrected for the pollution externality. This tax gap responds to the dynamic evolution of types, via the democratic process, and feeds back to the evolution of types, via an increasing or decreasing fitness of being a materialist. The model thus suggests that we should find a positive correlation between the shares of environmentalists and the strictness of environmental policy, which reflects the two-way causal link between values and policy.

6 Welfare Implications

Undertaking welfare analysis with changing preferences is well-known to be challenging. However, it also raises some interesting issues. Can we really say that a society comprising more or less of one type in the population is better off in a well-defined sense? This section explores that question.

A welfarist approach In a conventional model, we could define a social welfare function as a function of the utilities of both types. We work with a class of additive social welfare functions where

$$W\left(u^e, u^m, \mu\right) = \mu\omega(u^e(\mu)) + (1-\mu)\,\omega\left(u^m(\mu)\right), \tag{13}$$

where $\omega\left(\cdot\right)$ is an increasing concave function. If $\omega\left(\cdot\right)$ were linear, we would have a Utilitarian welfare function. To work with this welfare objective, we need to assume that payoffs are measurable and comparable. However, we have already implicitly assumed this in our formulation based on $\Delta\left(\mu\right)$ which implicitly assumes that citizens are capable of making such judgements when comparing the payoffs of different types.

The welfare analysis in our model is interesting, in that the standard approach to environmental policy would simply say that the optimal policy is to tax pollution at the Pigouvian level ($t = \lambda$) and that welfare cannot feasibly be any higher than at that tax rate. But once we allow for the possibility that the fractions of types are endogenous, this is no longer correct. We have to ask whether a society of environmentalists is happier if the environmental externality is completely eliminated in the long run, and not just mitigated via taxation.

Comparing the two steady states is straightforward. Exploiting (2) and (13), we have

$$W\left(u^e, u^m, 1\right) = \omega\left(\chi + y\right)$$

and

$$W\left(u^e, u^m, 0\right) = \omega(v(\hat{t}\left(0\right)) + y).$$

These expressions imply:

Proposition 3 *Welfare comparisons between the two steady states depend on parameter values:*

1. *If $\alpha < 1 + \chi$, then welfare is always higher with $\mu = 1$*

2. *If $\alpha \geq 1+\chi$, then there is a threshold value of λ such that welfare is highest with $\mu = 1$ – i.e., $\chi > v\left(\hat{t}\left(0\right)\right)$ for all λ above this threshold.*

Proof. If $\alpha < 1+\chi$ steady-state utility in the all-environmentalists steady-state is always higher than that in the all-materialists steady state as this implies $\omega\left(\chi+y\right) > \omega(v(\hat{t}\left(0\right))+y)$. What about the case when $\alpha > 1+\chi$? Suppose that $\lambda = 0$ (i.e., there is no pollution and therefore no corrective tax even when $\mu = 0$). Then, (1) implies that $v\left(\widehat{t}\left(0\right)\right) = \alpha - 1 > \chi$. But since $\lim_{\lambda\to\infty} v\left(\widehat{t}\left(\mu\right)\right) \to -\infty$, for all $\mu \in [0,1]$, it follows that for large enough λ, the consumption utility of materialists is always lower than the warm-glow utility of environmentalists. Hence for all $\alpha > 1$, there exists a value of λ for which being an environmentalist yields long-run higher utility than being a materialist. ■

Discussion This proposition makes intuitive sense. In its first case, the warm-glow allure of environmentalism is so strong that welfare is higher in a population consisting only of environmentalists. The second case is perhaps more interesting. It says that when $\lambda = 0$, materialism yields higher utility, and there is no need for a corrective tax But when λ is higher, and reaches a certain level, high taxation is needed even in a population of materialists. This means that their welfare is lower compared to the welfare level in an all-environmentalist population, which does not consume polluting good c even if the warm glow from environmentalism is very small.

The combination of Propositions 2 and 3 says that, whichever steady state is long-run optimal, there is no guarantee that society will converge to it. If $\alpha < 1+\chi$ then beginning with $\mu_0 < \hat{\mu}$ a society will converge to $\mu = 0$ which is sub-optimal. The same is true when $\alpha > 1+\chi$, provided λ is large enough. Convergence to environmentalism in our model requires a mass of environmentalists above a critical tipping point and without other forces supporting environmentalism, this will not happen.

One aspect of these results is that preference parameters play a role. This is an inevitable feature of any utility-based comparison of welfare across types. But regardless of the specific parametrization, the results also reflect a general issue in welfare-based models with culture dynamics. The welfare comparisons are based on steady-state utility *levels*, whereas cultural evolution hinge on utility *differences* between types. In our example, this leads to a bias against developing a welfare-improving environmentalist culture if societies begin with a low share μ of such types.

The welfare result is only reinforced if citizens have biased beliefs, or do not value pollution for other reasons. If citizens misperceive the true value of λ – failing, e.g., to internalize expert opinions on climate change – this gives an additional reason to believe that convergence to $\mu = 0$ is sub-optimal.

A failure of democratic politics? It is well-known that elections need not deliver welfare-maximizing outcomes. The classic example if the tyranny of the majority and the need to protect minority rights through courts and constitutional provisions. However, in this case the possibility of a sub-optimal outcome is due to a lack of commitment. Suppose that there would be a way to commit to the tax rate $\widehat{t}(1)$. This would encounter popular resistance if $\mu_s < 1$ – i.e., it would not be a political optimum. If credible. however, this prohibitive tax rate would influence the direction of the value dynamics, which by (10) would now be governed by:

$$\Delta\left(\mu_{s+1}\right) = \chi\varphi\left(\mu_{s+1}\right) - v\left(\widehat{t}(1)\right) = \chi\varphi\left(\mu_{s+1}\right) + \alpha - 1 > 0.$$

This implies

Proposition 4 *If it is possible to commit to* $\widehat{t}(1)$*, the society will converge* $\mu = 1$.

The key point here is that (expected) policy not only influences current payoffs. It also influences the dynamics of values and hence indirectly influences future welfare levels. This is a version of the failure of the political Coase theorem outlined in Acemoglu (2003). As discussed in a similar context – with endogenous manager types rather than endogenous consumer types – in Besley and Persson (2018), the problem is *not* that the decision-makers have a short horizon. Our earlier result on equilibrium policies would result even if politicians did internalize the future. Without the ability to commit to future policies, politicians as well as everybody else in society at s must take next period's equilibrium policy – that will depend on μ_{s+1} – as given. And, as we have seen, μ_{s+1} is itself based on future expected policy and hence not affected by current policy. However, this may no longer be true if the model had another state variable, such as the state of the environment.

These observations have more general resonance for thinking about optimal environmental policy when culture is changing. A forward-looking policy maker who anticipates cultural change may want to commit to a policy,

which is more draconian than would be justified by current preferences. This will mean lower current welfare but higher long-run welfare, as long there is little discounting of the future. Such a policymaker would act "as if" she discounted future generations' payoffs at a lower rate.

Remedies and enforcement Many (e.g., Stern 2015) have argued that, if a society believes that welfare is higher with environmentalism in the long run, it needs to implement stricter environmental policies than those consistent with the current Pigouvian optimum and the current political equilibrium.[15] Our analysis gives a normative and non-Paternalistic justification for non-majoritarian policy in an environmental context. The democratic policy process responds to short-term preferences, but this is no guarantee that society will converge to a long-run welfare optimum. If the environmental damages in our model were based on the cumulated stocks rather than the current flows of pollution – the right assumption in the case of fossil-fuel emissions and climate change – this point would be further reinforced.

Let us finally speculate about possible remedies in view of the results in Propositions 3 and 4. One way to think about them would be that they may justify a role for international organizations such as the EU. Such an organization could encourage policies which are not political equilibria for every country, as member countries with high levels of environmentalism would create a positive externality by pushing up environmental taxation. Another way to think about implementation of a better long-run equilibrium would be to invoke a role for environmental lobby groups. These would be pushing policy away from the conventional Pigouvian optimum. Via a kind of second-best logic, however, this may move the political equilibrium in a desirable direction, once the impact of changing values is taken into account.[16] Yet, another avenue to implementation might run through the judicial processes. Courts could adjudicate in favor stricter environmental policies, e.g., if future generations were given rights over current policies. But this would require that politicians who legislate those rights understood that environmentalism is endogenous, and that this legislation could not be

[15] Matauch et al (2018) develop a model of endogenous preferences in the context of Pigouvian taxation to support this conclusion.

[16] Interestingly, one of the goals of the Climate Leadership Council (CLC) – a club of private companies, including the oil giants – is to introduce a $40-a-ton fee on carbon-dioxide emissions in return for removing current climate change regulations and protecting companies from federal and state tort liability for historic emissions. (Guardian 2019)

repealed for short-run political gain.

7 Extensions and Open Issues

Our paper is very much a first pass and, at best, marks the beginning rather than the end of a research agenda. This section discusses a few possible extensions.

In our framework, an individual is an environmentalist mainly to convey to others through her consumption decisions that she cares about a certain cause, rather than because of a realistic expectation that she will make a difference by herself. We have not allowed environmentalists to influence social activism or political behavior outside of the voting context. Social movements and pressure groups may enhance the collective voice of environmentalists. In a standard setting, this might move policy (increase $\hat{t}(\mu)$ in our simple model). But our approach has also stressed the potential how altered policies might increase the cultural fitness of the environmental movement and change its numbers (raise future values of μ). In Besley and Persson (2019a) we look at social movements among nationalists which enter endogenously and enhance the salience of nationalism. The insights from that paper could be married with those from this paper.

We have also maintained a fixed party structure. However, the emergence of Green parties that seek direct policy influence, particularly by exploiting the coalition structures of proportional representation, may enhance the power of environmentalists and give them further power over policy (increase $\hat{t}(\mu)$ further). This will also have dynamic consequences if it affects the attractiveness of becoming an environmentalist. This could be modeled using the same approach as Besley and Persson (2019a) who look at endogenous nationalist party entry.

Another interesting extension would be direct socialization through the education system. We have already seen a link between education and environmental values in the WVS data. This link might reflect a general human-capital effect of reading more about the adverse consequences of human life styles for the planet. But, of course, governments may aim publicly-funded education on changing values. This is something that could be exploited in both directions (e.g., raising or lowering μ). In this context, there could be a role for forward-looking strategic policy making by government. Normative analysis would be politically controversial, but a positive analysis

would point to the same kind of political constraints as the choice of environmental regulation. In a similar vein, free and independent media as well as government-controlled media could influence values directly through their reporting of issues.

Finally, we have worked with a static model of society. However, there many features of the environment — not the least when it comes to climate change – are inherently dynamic. Modeling the interaction of a changing environment and the formation of changing values would be a challenging but important task.

8 Concluding Comments

At the root of this paper is the obvious point that any kind of environmental policy in a democratic society is constrained by what the citizens want. This has been vividly illustrated by recent real-world events. When, in 2018, French President Emmanuel Macron tried to raise the tax on motor fuels – a move that many would describe as environmentally sound – the Gilets Jaunes took to the streets of Paris to protest. U.S. President Donald Trump's recent decision to withdraw from the Paris climate-change agreement was very popular among his supporters. Many well-meaning people who are environmentalists would advocate bans on polluting emissions (corresponding to the prohibitive tax $\widehat{t}(1)$ in our model) but ignore the fact that this is far from a political equilibrium. These political constraints have both static and dynamic consequences.

The climate change politics in our framework moves beyond standard "Pigouvian" models of policy-making which dominate the literature by building a role for changing preferences. This is timely given the current dynamics of social movements – such as the UK "Extinction Rebellion" and the protests among young people in Europe started by Greta Thunberg's school strike – which aim at creating behavioral and policy change. We have illustrated an interesting interplay between affecting behavior in traditional ways (via economic incentives) and the influence of policy on culture. Social movements stress the importance of declaring a "climate emergency," which is often dismissed as an empty gesture of virtue signalling. But our analysis shows why virtue signalling can be a driver of cultural change. If environmentalists were just miserable about pollution and climate damages and took actions which reduced their own material living standards, environmentalism as a social

movement may not catch on. It is the positive message of environmentalism as a virtue (a higher χ in our model) that can create an environmentalist culture. Declaring a common cause can increase the perceived virtue of private actions. Exploring such issues further opens up a rich potential agenda on the political economics of climate policy.

We have used our model to explore a new issue in political economics – namely the interplay between democratic politics and endogenous environmental values. In our setting, policy not only shapes current welfare outcomes but expected policy also influences future values via cultural evolution. By responding to citizens' preferences, politics can create a kind of momentum that can drive multiple steady-states. There is no reason to believe that society will converge to the long-run outcome with the highest welfare level. Moreover, this happens even when the political equilibrium is choosing something close to a Pigouvian optimum based on current preferences. In our model, political preferences are not distorted away from standard welfare objectives – in fact, the probabilistic-voting framework produces a Utilitarian policy outcome.

Although our application is specific, we believe it delivers some takeaway messages of wider significance.

First, policy choices can affect the socialization of types by affecting their cultural (psychological) fitness. Even though environmentalism is a natural application, we believe that this insight fits many other contexts.

Second, with endogenous values one must consider how social welfare depends upon the composition of population types. But then one has to grapple with the thorny issue whether citizens in some societies have "better values" than others. Our paper has suggested a new way of looking at the welfare economics of environmental taxation. In particular, we show why one may not want to succumb to the usual Pigouvian logic that optimal policy should reflect only current preferences. If society's preferences are themselves endogenous, then long-run desirable policies may be a lot more draconian. However, democratic societies would find it very hard to bring such draconian policies about. That some of today's citizens ignore environmental degradation does not make the problem go away, and the experienced utility of living in a damaged environment may eventually come home to roost in a variety of ways. A similar logic may apply in other policy spheres.

Third, there is no reason to believe that a cultural evolutionary process will converge to a social optimum. A system where *relative,* rather than absolute, payoffs drive cultural dynamics will almost always deliver such a

conclusion. This is further compounded when politics ensures that current preferences drive policy choices.

Fourth, our framework has highlighted how a political process, where policy is made by current majorities not only affects current outcomes but also emerging values. If we assume – as did Acemoglu and Robinson (2000) – that it is easier to commit to future institutions for policymaking than to future policies, our results suggest that it may be desirable to find institutional frameworks which reduce the responsiveness of environmental policy to current preferences. This may seem to run against one of the assumed virtues of democracy, to deliver policy outcomes that respond to the wishes of the current majority. Yet, many societies routinely delegate policy choices to more far-sighted institutions in other domains, such as central banking.

Finally, economists have been reluctant to embrace cultural dynamics in their analyses. However, our modeling suggests that such reluctance could neglect an important aspect of policy-making. Some may find it unpalatable to say that we have to change people's values to fundamentally change the world. But as we have shown, thinking about values is a complement to the conventional approach to optimal policy choices. More generally, our analysis suggests that failing to consider how social and cultural values change in response to policy may give an incomplete account of human progress.

References

[1] Acemoglu, Daron, [2003], "Why Not a Political Coase Theorem?: Social Conflict, Commitment, and Politics," *Journal of Comparative Economics*, 31, 620–52.

[2] Acemoglu, Daron and James Robinson, [2000], "Why Did the West Increase the Franchise?

[3] Acemoglu, Daron and James Robinson, [2006], "Economic Backwardness in Political Perspective," *American Political Science Review*, 100, 115-131.

[4] Aghion, Philippe, Roland Benabou, Ralf Martin and Alexandra Roulet, [2019], "Environmental Preferences and Technological Choices: Is Market Competition Clean or Dirty?" unpublished manuscript.

[5] Aghion, Philippe and Patrick Bolton, [1990], "Government Debt and the Risk of Default: A Politico-Economic Model of the Strategic Role of Debt," in Rudiger Dornbusch and Mario Draghi (eds), *Public Debt Management: Theory and History*, Cambridge University Press: Cambridge, England.

[6] Akerlof, George and Rachel Kranton, [2000], "Economics and Identity," *Quarterly Journal of Economics*, 115, 715–753.

[7] Akerlof, George, and Rachel Kranton, [2010], *Identity Economics: How Our Identities Shape Our Work, Wages, and Well-Being*, Princeton: Princeton University Press.

[8] Almond, Gabriel and Sidney Verba, [1963], *The Civic Culture: Political Attitudes and Democracy in Five Nations*, New York: Sage Publications.

[9] Andreoni, James, [2006], "Philanthropy", in Serge-Christophe Kolm and Jean Mercier Ythier, Ed(s), *Handbook of the Economics of Giving, Altruism and Reciprocity*, Elsevier, 2006, Volume 2, Chapter 18, Pages 1201-1269.

[10] Benabou, Roland and Jean Tirole, [2006], "Incentives and Prosocial Behavior," *American Economic Review*, 96(5), 1652-1678.

[11] Besley, Timothy, [2017], "Aspirations and the Political Economy of Inequality," The 2016 Hicks Lecture in Economics, *Oxford Economic Papers*, 69, 1–35.

[12] Besley, Timothy, [2019], "State Capacity, Reciprocity and the Social Contract," forthcoming in *Econometrica*.

[13] Besley, Timothy and Stephen Coate, [1997], "An Economic Model of Representative Democracy", *Quarterly Journal of Economics*, 112, 85-114.

[14] Besley, Timothy and Stephen Coate, [1998], "Sources of Inefficiency in a Representative Democracy: A Dynamic Analysis", *American Economic Review*, 88, 139-156.

[15] Besley, Timothy and Maitreesh Ghatak, [2005], "Competition and Incentives with Motivated Agents," *American Economic Review*, 95(3), 616-636.

[16] Besley Timothy and Torsten Persson, [2009], "The Origins of State Capacity: Property Rights, Taxation and Politics," *American Economic Review*, 99(4), 1218-1244.

[17] Besley Timothy and Torsten Persson, [2011], *Pillars of Prosperity: The Political Economics of Development Clusters*, Princeton, NJ: Princeton University Press.

[18] Besley Timothy and Torsten Persson, [2018] "Organizational Dynamics: Culture, Design, and Performance," mimeographed manuscript.

[19] Besley Timothy and Torsten Persson, [2019a], "The Rise of Identity Politics," mimeographed manuscript.

[20] Besley Timothy and Torsten Persson, [2019b], "Democratic Values and Institutions," *American Economic Review: Insights*, forthcoming.

[21] Bezin, Emeline, [2015], "A Cultural Model of Private Provision and the Environment," *Journal of Environmental Economics and Management*, 71, 109–124.

[22] Bisin, Alberto and Thierry Verdier, [2011], "The Economics of Cultural Transmission and Socialization," Chapter 9 in Benhabib, Jess (ed.), *Handbook of Social Economics*, Volume 1A, Elsevier.

[23] Bisin, Alberto and Thierry Verdier, [2017], "On the Joint Evolution of Culture and Institutions," NBER Working Paper No. 23375

[24] Bowles, Samuel, [1998], "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions," *Journal of Economic Literature*, 36, 75-111.

[25] Bowles, Samuel, Jung-Kyoo Choi, and Astrid Hopfensitz, [2003], "The Co-evolution of Individual Behaviors and Social Institutions," *Journal of Theoretical Biology*, 223, 135–147.

[26] Boyd, Robert and Peter Richerson, [1985], *Culture and the Evolutionary Process*, Chicago, IL: University of Chicago Press.

[27] Cavalli-Sforza, Luigi L. and Marcus Feldman, [1981], *Cultural Transmission and Evolution*, Princeton, NJ: Princeton University Press.

[28] Coughlin, Peter, [1992], *Probabilistic Voting Theory*, Cambridge: Cambridge University Press.

[29] Dekel, Eddie, Jeffrey Ely and O. Yilankaya, [2007], "Evolution of Preferences," *Review of Economic Studies*, 74, 685–704.

[30] Dietz, Thomas, Amy Fitzgerald, and Rachael Shwom, [2005], "Environmental Values," *Annual Review of Environment and Resources*, 30, 335–72.

[31] Downs, Anthony, [1957], *An Economic Theory of Democracy*, New York: Harper Collins.

[32] Druckman, James N. and Arthur Lupia, [2016], "Preference Change in Competitive Political Environments," *Annual Review of Political Science*, 19, 13–31.

[33] Guardian (2019) "Microsoft Joins Group Seeking to Avoid Historic Climate Change Lawsuit," https://www.theguardian.com/technology/2019/may/01/microsoft-joins-group-seeking-to-avoid-climate-change-lawsuit

[34] Güth, Werner and Menahem E. Yaari, [1992], "Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach," in Witt, Ulrich (ed.) *Explaining Process and Change: Approaches to Evolutionary Economics*, Ann Arbor, MI: University of Michigan Press.

[35] Inglehart, Ronald, Miguel Basanez, Jaime Diez-Medrano, Loek Halman, and Ruud Luijkx (eds.), [2004], *Human Beliefs and Values: A Cross-cultural Sourcebook Based on the 1999–2002 Values Surveys*, Mexico City: Siglo XXI

[36] Krusell, Per, and Jose-Victor Rios-Rull, [1996], "Vested Interests in a Theory of Growth and Stagnation," *Review of Economic Studies*, 63, 301–29.

[37] Krusell, Per, Quadrini, Vincenzo and Jose-Victor Rios-Rull, [1997], "Politico-Economic Equilibrium and Economic Growth," *Journal of Economic Dynamics and Control,* 21, 243-272

[38] Lindbeck, Assar and Jörgen Weibull, (1987). "Balanced Budget Redistribution as the Outcome of Political Competition," *Public Choice,* 52, 273-297.

[39] Lorenzoni, Irene and Nick Pidgeon, [2006], "Public Views on Climate Change: European and USA Perspectives," *Climatic Change*, 77, 73-95.

[40] McAdam, Doug, [2017], "Social Movement Theory and the Prospects for Climate Change Activism in the United States," *Annual Review of Political Science*, 20, 189–208.

[41] Mattauch, Linus, Cameron Hepburn, and Nicholas Stern, [2018], "Pigou Pushes Preferences: Decarbonisation and Endogenous Values," Grantham Research Institute on Climate Change and the Environment, Working Paper No. 314.

[42] Meltzler, Alan and Scott Richards, [1981], "A Rational Theory of the Size of Government," *Journal of Political Economy*, 89, 914-927.

[43] Milesi-Ferretti, Gian-Maria and Enrico Spolaore, [1994], "How Cynical Can an Incumbent Be? Strategic Policy in a Model of Government Spending, " *Journal of Public Economics*, 55, 121-40.

[44] Nordhaus, William, [2008] *A Question of Balance: Weighing the Options on Global Warming Policies*, New Haven and London: Yale University Press.

[45] Oates, Wallace and Robert Portney, [2003], "The Political Economy of Environmental Policy," in *Handbook of Environmental Economics*, Volume 1, Edited by K-G. Maler and J.R. Vincent, Elsevier.

[46] Osborne, Martin and Al Slivinski, [1996], "A Model of Political Competition with Citizen-candidates", *Quarterly Journal of Economics*, 111, 65-96.

[47] Persson, Torsten, Gérard Roland, and Guido Tabellini, [1997], "Separation of Powers and Political Accountability," *Quarterly Journal of Economics*, 112, 1163–1202.

[48] Persson, Torsten and Guido Tabellini, [2000], *Political Economics: Explaining Economic Policy*, Cambridge, MA: MIT Press.

[49] Persson, Torsten and Lars Svensson, [1989], "Why a Stubborn Conservative Would Run A Deficit: Policy with Time-Inconsistent Preferences," *Quarterly Journal of Economics*, 104, 325-346.

[50] Rogoff, Kenneth, [1985], "The Optimal Degree of Commitment to an Intermediate Target," *Quarterly Journal of Economics,* 100, 1169-90.

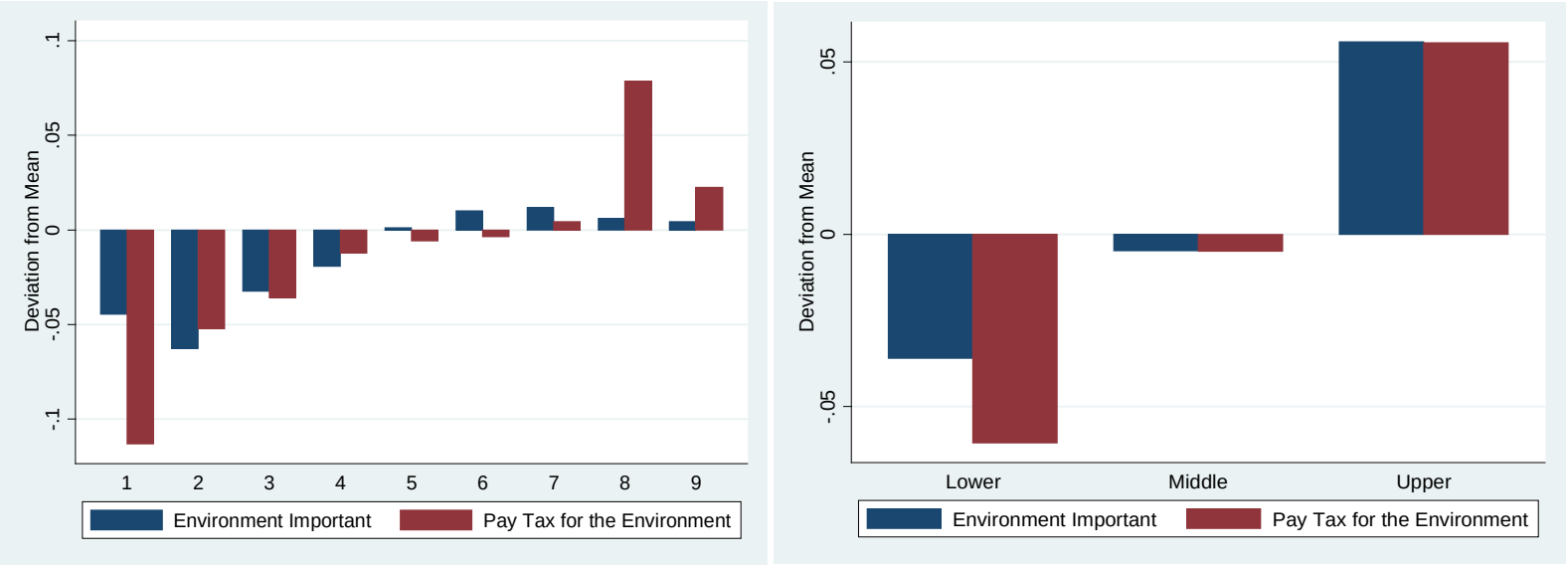
[51] Sandholm, William H., [2010], *Population Games and Evolutionary Dynamics*, Cambridge, MA: MIT Press.

[52] Stern, Nicholas H., [2015], *Why Are We Waiting? The Logic, Urgency, and Promise of Tackling Climate Change*, Cambridge MA: MIT Press.

[53] Tabellini, Guido and Alberto Alesina, [1990], "Voting on the Budget Deficit," *American Economic Review*, 80, 37-49.

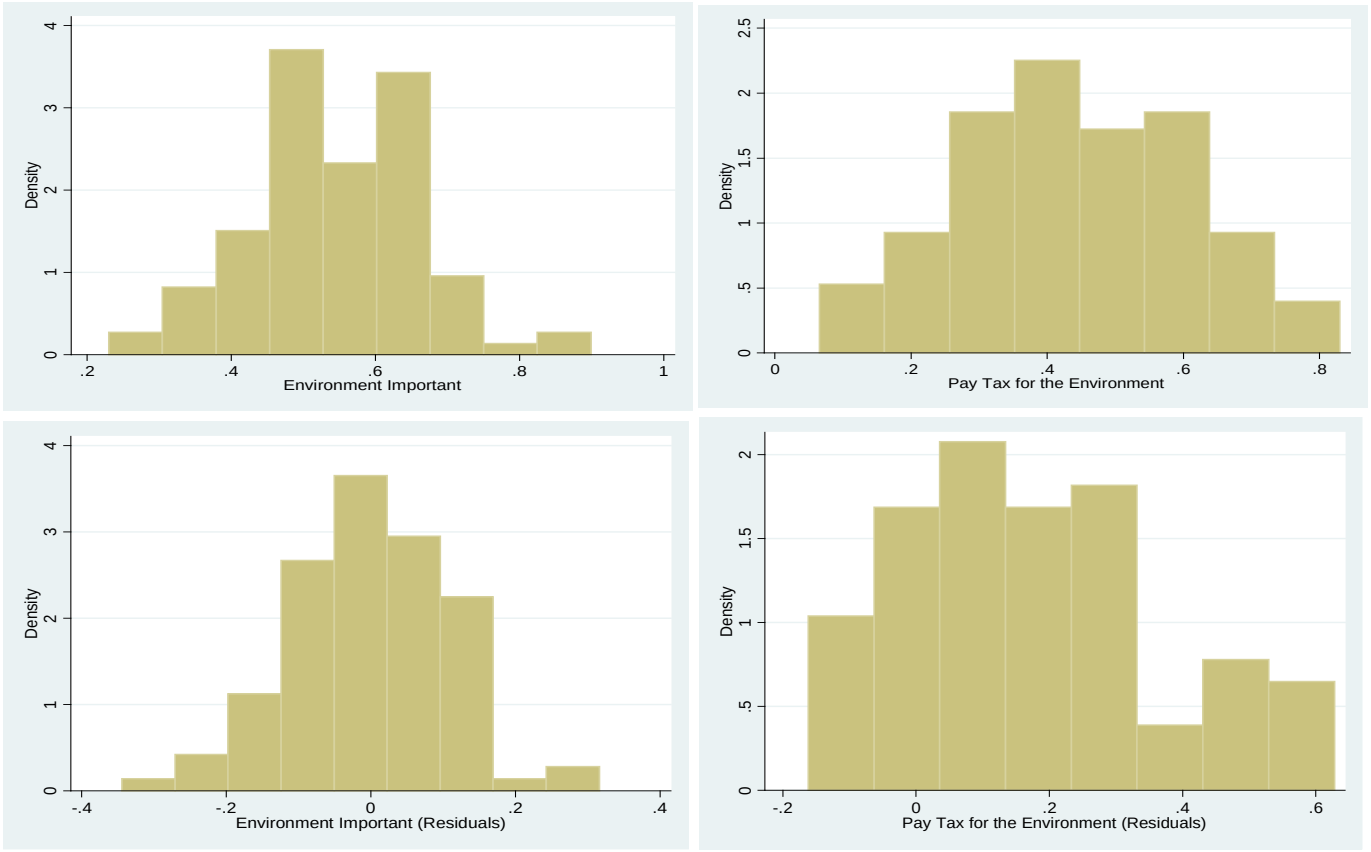
[54] Tsebelis, George, [2002], *Veto Players: How Political Institutions Work*, New York and Princeton, NJ: Russell Sage Foundation and Princeton University Press.

Figure 1: Environmental Attitudes by Birth-decade and Education



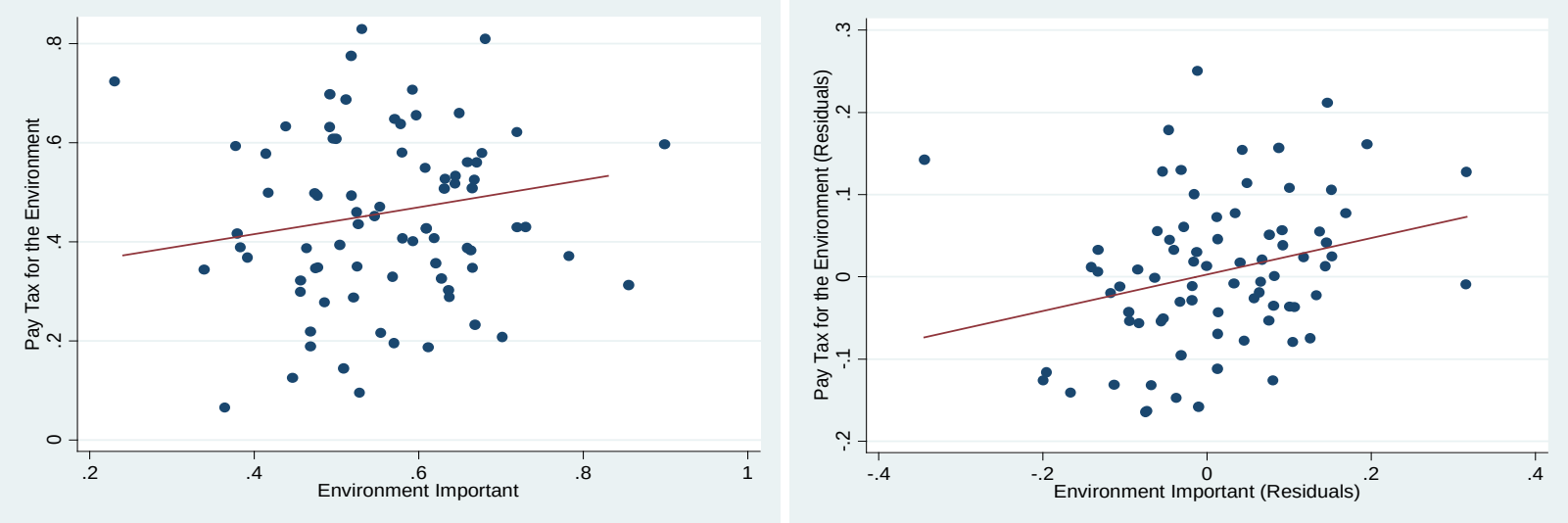
Notes: The graphs show deviations from means for different groups using data from answers to two WVS questions, one where respondents are asked whether the environment is an important policy priority (question B002 in the last WVS wave), and the second whether they “strongly agree” or “agree” with an "increase in taxes if used to prevent environmental pollution" (question B008). The left graph shows average deviations from overall country means among respondents who belong to 10-year cohorts born in the 1910s and onwards, while the right graph shows average deviations from overall country means among respondents in three groups according to their level of education.

Figure 2: Cross-country Variation in Environmental Attitudes



Notes: Each histogram shows the variation across countries in the share of respondents who think the environment is a priority (left two graphs), and whether they support taxes to help the environment (right two graphs) – see the text and Note to Figure 1. The top row shows the average raw share, while the bottom row shows the average share adjusted for individual characteristics. The latter is based on a linear regression at the WVS individual level with an individual dummy on the LHS, and a dummy for gender, ten dummies for income groups, three for education groups, three age bands, and WVS wave dummies on the RHS.

Figure 3: Country-level Correlation Between Environmentalism and Support for Environmental Taxes



Notes: Both graphs show the correlation in the WVS between country means for holding environmentalist values and for supporting taxes to help the environment. The left graph shows the raw data, while the right graphs show the mean country residuals adjusted for individual characteristics. To define these residuals, we run a linear regression at the WVS individual level with an individual dummy on the LHS, and a dummy for gender, ten dummies for income groups, three for education groups, three age bands, and WVS wave dummies on the RHS.