Compliance without Conviction : Carbon Pricing and the Dynamics of Environmental Values*

Can Askan Mavi† Nouhoum Touré‡

October 20, 2026

Abstract

Carbon pricing not only changes behavior; it can also shape the environmental values that sustain climate policy. We present a dynamic political-economy model in which environmental values, carbon-intensive behavior, and carbon taxes evolve jointly. A greener electorate chooses a higher carbon tax. Once environmentalists have exhausted their adjustment possibilities, further tax-induced abatement is supplied by materialists. Their behavior converges toward that of environmentalists even though their underlying values remain unchanged. This convergence reduces the extent to which environmental values translate into distinct behavior and weakens their transmission, even as their relative payoff continues to rise. We show that this feedback can prevent environmental values from spreading throughout the population and can make a carbon tax that is efficient for the current generation dynamically excessive. We also show that policies generating the same current abatement can have opposite cultural effects. Indeed, raising the carbon tax compresses behavioral differences and weakens the transmission of environmental values, whereas expanding low-carbon alternatives widens these differences and strengthens their transmission. More generally, the long-run effects of policy depend not only on how much behavior changes, but also on whose behavior changes.

Keywords: Carbon pricing; environmental values; cultural transmission; behavioral incidence; political economy.

JEL: Q58, H23, D72, D91, Z13.

*We are grateful to Ablam Estel Apeti, Thomas Baudin, Rémi Bazillier, Ibrahim Ben Araar, Emeline Bezin, Stefano Bosi, Raouf Boucekkine, David de la Croix, Giorgio Fabbri, Victor Hiller, Tanguy Le Fur, Paul Maarek, Fabio Mariani, Paolo Melindi-Ghidi, Grégory Ponthière, Fabien Prieur, Natacha Raffin, Ariane Salem, Avner Seror, Josselin Thuilliez, Paola Villar, Thierry Verdier, and Leonard Wantchekon for helpful discussions and comments. All remaining errors are our own.

†Université Paris-Saclay, INRAE, AgroParisTech, Paris-Saclay Applied Economics, 91120 Palaiseau, France. can-askan.mavi@inrae.fr

‡Université Paris 1 Panthéon-Sorbonne, Centre d'Économie de la Sorbonne, UMR 8174 CNRS, Paris, France. nouhoum.toure@univ-paris1.fr

1. Introduction

Carbon pricing has become a central instrument of climate policy, and now covers just over 29 percent of global emissions (World Bank, 2026). Yet, its political support can prove fragile once it reaches the fuels that households use for transportation and heating, as illustrated by Canada's removal of its consumer carbon price in 2025 and the European Union's decision to postpone to 2028 the launch of its new Emissions Trading System for road transport and buildings (ETS2).[1] These episodes have brought the political sustainability of carbon pricing to the center of current policy debates.

Existing work on carbon pricing has developed along two complementary dimensions. One strand studies its immediate effects on abatement and on the distribution of its costs across households (Andersson, 2019; Klenert et al., 2018). Another examines its longer-run interaction with environmental values. A growing body of work shows that these values are not fixed, but evolve with the economic and political environment. As emphasized by Besley and Persson (2019b), Bezin (2019), and Besley and Persson (2023), greener policies and technologies can make green values more attractive and thereby reinforce their diffusion. We connect these two dimensions by studying how the identity of the households that adjust to climate policy shapes the transmission of environmental values. We refer to this margin as the *behavioral incidence* of climate policy. When environmental values are partly transmitted through observed behavior (Boyd and Richerson, 1985; Bezin, 2015), behavioral incidence determines whether climate policy preserves or compresses the behavioral differences through which those values are reproduced across generations.

The intergenerational transmission of values may depend not only on what parents believe or communicate, but also on the behaviors their children observe (Boyd and Richerson, 1985; Henrich, 2009). This mechanism is particularly relevant for environmental values. Using parent–child data from Danish households, Grønhøj and Thøgersen (2009) document intergenerational similarities in environmental attitudes and practices, while Grønhøj and Thøgersen (2012) show that family norms are especially strongly associated with adolescents' pro-environmental behavior when reflected in their parents' actions. Evidence from China reinforces the importance of observability: Jia and Yu (2021) find that the association between parents' and children's pro-environmental behavior is substantially weaker when parental actions are not directly observed by children. Importantly, however, environmental policies may change observed behavior without necessarily reinforcing the underlying values. For instance, Xu et al. (2023) find that financial incentives increased household recycling while reducing intrinsic pro-environmental motivation. Taken together, these findings suggest that policy-induced changes in behavior may affect the channels through which environmental values are transmitted across generations. Yet this feedback remains largely absent from existing models of environmental policy and cultural dynamics, which typically emphasize how policies and technologies affect the relative attractiveness of environmental values (Besley and Persson, 2023; Bezin, 2019), rather than how policy changes the behavioral differences through which those values are expressed and transmitted. We study this missing margin and ask how policy-induced behavioral convergence feeds back into cultural transmission and, in turn, shapes the long-run sustainability of climate policy.

To do so, we provide a dynamic political-economy model in which environmental values, carbon-

[1] In July 2026, ten member states asked the European Commission to reconsider ETS2 (Abnett, 2026). More recently, the surge in oil prices caused by the war in Iran led several governments, from Spain to Pakistan, to cut fuel taxes. These events recall the Yellow Vests movement, which led the French government to freeze the planned increase of its carbon tax in 2018 (see Douenne and Fabre, 2022).

intensive behavior, and climate policy evolve jointly. We consider a population composed of two cultural groups, *environmentalists* and *materialists*. Environmentalists attach a higher value to environmental quality and consume less of a carbon-intensive good at any given price. Households can reduce their carbon use only down to a common lower bound, which summarizes the technological and infrastructural constraints that limit substitution in the short run, such as housing and heating technologies, transportation options, or the energy mix (Hochmuth et al., 2026; Ascari et al., 2026).[2] Since environmentalists consume less carbon even in the absence of policy, they reach this lower bound at a lower carbon price than materialists.

Climate policy is determined through probabilistic voting between two office-motivated parties, with carbon-tax revenues returned to households. Since environmentalists attach a higher value to environmental quality, an increase in their population share raises the equilibrium carbon tax, which maximizes the average welfare of the current generation. Regarding cultural evolution, we build on the socialization mechanism proposed by Besley (2017), in which children are influenced both by their parent and by a role model drawn from society. In that framework, when the two role models hold different cultural traits, the probability of adopting one trait depends on its relative utility. We modify this mechanism by allowing children to learn about cultural traits from the behavior associated with them. The transmission advantage of environmental values therefore depends on the behavioral difference between environmentalists and materialists, rather than solely on the relative payoff associated with these values.[3]

These elements generate a two-way interaction between environmental values and climate policy. As the population becomes greener, political support for carbon pricing increases and the equilibrium carbon tax rises. As long as both cultural groups remain able to reduce their carbon use, they respond to the higher tax in the same way, so that carbon use falls while the behavioral difference between environmentalists and materialists remains unchanged. Once environmentalists have exhausted their possibilities to adjust, however, further tax-induced abatement must be supplied by materialists. Their behavior then converges toward that of environmentalists even though their underlying values remain unchanged, reducing the extent to which environmental values are expressed in distinctive behavior and thereby weakening their transmission. We refer to this mechanism as *compliance without conviction*: materialists increasingly adopt the behavior associated with environmentalism without adopting the values themselves. Hence, while a larger share of environmentalists always leads to a higher carbon tax, the effect of policy on cultural dynamics changes once environmentalists cease to supply the marginal adjustment.

A key implication of this mechanism is that the relative payoff of environmental values and their transmission advantage need not move in the same direction. Below the threshold at which environmentalists become constrained, a greener electorate raises the carbon tax and reduces carbon use without affecting the behavioral difference between environmentalists and materialists. Environ-

[2]The ability of households to adjust to carbon prices on household fuels is highly heterogeneous. Eisner et al. (2021) show that households' vulnerability to carbon pricing depends importantly on housing characteristics, location, and household composition. In the Netherlands, for instance, the national environmental assessment agency estimates that ETS2 will raise the monthly energy and fuel bills of large households by up to 70 euros by 2030, whereas households that have already switched to heat pumps and electric vehicles will avoid these costs.

[3]Our results do not rely on children being unable to observe the constraints that shape behavior. When these constraints are observable, behavioral convergence still weakens the transmission of environmental values, although cultural reversal then requires some prior skepticism about their behavioral content. We also obtain the same qualitative result when cultural transmission results from deliberate parental socialization, as in Bisin and Verdier (2001). In this case, parents understand the constraints behind observed behavior, but the return to transmitting environmental values falls when these values make less difference to the child's behavior, as in Carvalho (2013).

mental values therefore continue to spread. Once environmentalists become constrained, however, materialists progressively adjust their behavior toward that of environmentalists, and the transmission advantage of environmental values declines. Environmental values continue to spread for some time, but begin to decline when the carbon tax becomes sufficiently high.[4] The economy consequently converges to an interior steady state in which part of the population remains materialist, even though the relative payoff of environmentalists continues to increase with their population share. This result differs from Besley (2017) and Besley and Persson (2019b), where cultural dynamics are driven by the relative payoff associated with each trait. In our framework, cultural transmission also depends on the behavioral differences generated by these traits. Environmental values may therefore become relatively more attractive while becoming less widespread.

Our results also allow us to shed new light on the design of climate policy. We find that policies generating the same current reduction in carbon use may have very different long-run effects depending on which households adjust. To illustrate this result, consider an economy in which environmentalists have already exhausted their possibilities to reduce carbon use. A further increase in the carbon tax then induces materialists to adjust, bringing their behavior closer to that of environmentalists and weakening the transmission of environmental values. By contrast, expanding low-carbon alternatives allows environmentalists to reduce their carbon use further, increases the behavioral difference between the two cultural groups, and favors the transmission of environmental values. Hence, aggregate abatement is not sufficient to characterize the long-run effects of climate policy: the identity of the households supplying the adjustment also matters. We refer to this additional margin as the *behavioral incidence* of climate policy, in contrast to the distributional incidence of carbon pricing, which concerns how its economic costs are distributed across households and regions (Goulder et al., 2019; Conte et al., 2026; Bigio et al., 2026; Konradt and Mangiante, 2026). This distinction also explains why lump-sum compensation does not affect our cultural mechanism: transfers change households' resources but not their carbon use. We also show that this cultural feedback has important welfare implications. Once behavioral convergence begins, the equilibrium carbon tax may become dynamically excessive even though it remains optimal for the current generation. Indeed, around the static optimum, a small reduction in the tax has only a limited effect on current welfare, while preserving the behavioral differences that support the transmission of environmental values. This distortion can therefore arise even while environmental values are still spreading. A lower committed tax can then preserve their transmission and eventually lead to lower carbon use. For a sufficiently patient social planner, it also yields higher discounted welfare.

Although our model is too stylized to provide quantitative policy guidance, the results have implications for current debates on carbon pricing. Much of this debate concerns how households should be compensated for the costs of carbon taxation (Goulder et al., 2019; Douenne et al., 2026). In our framework, however, compensation and investments in low-carbon alternatives play very different roles. Transfers affect households' resources, whereas investments in building renovation, clean heating, or low-carbon mobility expand their possibilities to adjust. The latter therefore affect which households reduce carbon use and, through this channel, the transmission of environmental values. This distinction also has a welfare implication. Once the most motivated households have exhausted their adjustment possibilities, further increases in the carbon price may weaken the transmission of

[4]The lower bound is not the source of this reversal, but provides a transparent way of generating behavioral convergence. As discussed in Section 3, the same reversal arises whenever the equilibrium tax sufficiently compresses the behavioral difference between cultural groups.

environmental values. This does not imply that carbon prices should generally be low. By expanding households' adjustment possibilities, investments in low-carbon alternatives allow higher carbon prices to be sustained before behavioral convergence begins to weaken cultural transmission.

More generally, our mechanism applies beyond environmental policy whenever individuals with different preferences are induced to behave similarly and these preferences are transmitted through behavior. Material rewards, for instance, can increase blood donations among initially less motivated individuals (Goette and Stutzer, 2020), while financial incentives for recycling can weaken the link between pro-environmental motivation and observed behavior (Ling and Xu, 2021). Similarly, regulations affecting religious practices can alter the behavioral expression of religious values (Carvalho, 2013). In all these cases, the long-run effects of policy depend not only on its aggregate behavioral impact, but also on how this behavioral response is distributed across individuals with different preferences.

Our paper is most closely related to the literature on environmental values and the political economy of green transitions. Besley and Persson (2019b) show how environmental values and environmental policy can jointly evolve through political competition, while Besley and Persson (2023) study a green transition in which values, technologies, and policy coevolve over time. Bezin (2015) emphasizes the role of observed behavior in the intergenerational transmission of environmental preferences, whereas Bezin (2019) studies the interaction between environmental preferences, technological change, and environmental policy. Related contributions examine how endogenous preferences affect the design and effectiveness of climate policy (Mattauch et al., 2022; Torren-Peraire et al., 2025). Despite differences in their mechanisms, these papers share the idea that green values, technologies, and policies may reinforce one another over time. Greener values favor greener behavior and policies, which in turn make green values more attractive or facilitate their diffusion. Our contribution is to emphasize the role of behavior in sustaining this dynamic complementarity. In most of the literature, greener policies and technologies reinforce environmental values by increasing their relative attractiveness or facilitating their diffusion. We introduce an additional channel through which policy affects cultural transmission. Once environmentalists have exhausted their possibilities to adjust, a higher carbon tax mainly induces materialists to reduce their carbon use. Their behavior therefore becomes closer to that of environmentalists, even though their underlying preferences remain unchanged. As a result, environmental values become less distinctive in observed behavior and their transmission weakens. Importantly, this can occur while the relative payoff of environmental values continues to increase. Hence, the very success of environmental values in shaping policy may ultimately weaken their own transmission.

Our analysis also contributes to the broader literature on the joint evolution of culture and political institutions (Bisin and Verdier, 2024; Besley and Persson, 2019a; Touré, 2021). Bisin and Verdier (2024) provide a general framework for studying how cultural and institutional dynamics interact, while Besley and Persson (2019a) and Touré (2021) illustrate how these interactions can generate persistence and different long-run outcomes. Our contribution is to identify a particular mechanism through which the feedback from policy to culture becomes non-monotonic. In our framework, the same political rule can have very different cultural consequences depending on which individuals adjust their behavior in response to policy. As a result, stronger policy may initially leave cultural transmission unaffected and later weaken it, even though neither preferences nor the political determination of policy have changed.

Finally, our mechanism is related to work on the cultural effects of policy. In Carvalho (2013), the regulation of veiling changes the return to religious socialization and thereby affects the long-run prevalence of religious values. A related literature shows that policies targeting identities, languages, or socializing institutions may induce the targeted groups either to reinforce their identity (Fouka, 2020; Abdelgadir and Fouka, 2020; Squicciarini, 2020; Bazzi et al., 2026) or to assimilate (Fouka, 2019), while Carvalho et al. (2024) develop a theory of resistance to cultural transmission through education. Our framework differs from these contributions in an important way. In this literature, cultural change generally follows from policies that directly affect cultural practices or the institutions through which culture is transmitted. In our model, carbon pricing has no cultural objective and materialists do not resist it. They simply respond to the price signal, and it is precisely their compliance that weakens the transmission of environmental values. Moreover, the tax is itself determined by the cultural composition of the population. The policy brought about by the diffusion of environmental values can therefore weaken their own reproduction, even though it neither targets culture nor generates resistance.

The remainder of the paper is organized as follows. Section 2 presents the economic and political environment. Section 3 introduces cultural transmission and characterizes the joint dynamics of environmental values and climate policy. Section 4 studies the implications of behavioral incidence for the design of climate policy. Section 5 analyzes the intertemporal welfare consequences of the political equilibrium and compares it with a committed policy. Section 6 considers purposeful parental socialization, and Section 7 concludes.

2. The Model

Time is discrete and indexed by t. The economy is populated by overlapping generations that live for two periods. In each period, there is a unit mass of households, each consisting of one adult and one child. Adults make the economic and political decisions of the current period and are replaced by their children in the following period.

The adult population is divided into two cultural groups: *environmentalists*, denoted by e, and *materialists*, denoted by m. The two groups differ in their preferences. Environmentalists attach greater value to environmental quality and have a lower desired consumption of a carbon-intensive good than materialists. Let q_t denote the proportion of environmentalists in generation t, so that $1 - q_t$ is the proportion of materialists. Culture is represented by the distribution of environmental preferences in the population. Climate policy is represented by a carbon tax determined through electoral competition between two office-motivated parties. Given the prevailing policy, adults choose their consumption of the carbon-intensive good subject to a minimum carbon requirement. Aggregate carbon use determines environmental quality, while carbon-tax revenues are returned to households through a uniform lump-sum transfer. Environmental preferences are transmitted across generations through family exposure and social learning from adult role models. Within a period t (q_t being given), events unfold as follows:

1. Two office-motivated parties compete by proposing a carbon tax τ_t.

2. Given the carbon tax, adults choose their consumption of the carbon-intensive good. Aggregate carbon use determines environmental quality and carbon-tax revenues are rebated to households.

As usual the model is solved by backward induction: household behavior is characterized first for a given carbon tax, and the resulting indirect utilities determine the political equilibrium. Once the current equilibrium has been determined, environmental preferences are transmitted across generations through family exposure and social learning from adult role models. This process determines the environmentalist share q_{t+1} in the next generation. Household behavior is characterized in Section 2.1, the political equilibrium in Section 2.2, and equilibrium payoffs in Section 2.3. Cultural transmission and the resulting dynamics of q_t are studied separately in Section 3.

2.1. Household Behavior

Adults differ in their cultural type, indexed by $i \in \{e, m\}$, where e denotes environmentalists and m materialists. Each adult consumes a numeraire good, c_t^i, and a carbon-intensive good, x_t^i. The latter should be understood broadly as consumption associated with transportation, heating, energy use, or other activities whose carbon content cannot be eliminated immediately. Let Q_t denote environmental quality. The flow utility of an adult of type i is [5]

$$u_t^i = c_t^i + x_t^i - \frac{\rho}{2}\left(x_t^i - \alpha_i\right)^2 + \sigma_i Q_t, \qquad \rho > 0. \tag{1}$$

The parameter $\alpha_i > 0$ represents a type-specific reference level for carbon-intensive consumption. Given the normalization of the pre-tax price to one, it coincides with the level of carbon-intensive consumption chosen by type i in the absence of carbon taxation, so that $x_i^*(0) = \alpha_i$. A larger α_i therefore corresponds to a higher baseline level of carbon-intensive consumption. The parameter $\sigma_i \geq 0$ measures the value that type i attaches to environmental quality. We interpret differences in α_i and σ_i as two manifestations of the same underlying cultural orientation, consistent with models in which environmental types differ both in their valuation of environmental quality and in the behavior induced by environmental preferences (Besley and Persson, 2019b; Bezin, 2015). Environmentalists have a lower reference level for carbon-intensive consumption and attach greater value to environmental quality than materialists. The parameter $\rho > 0$ captures the private cost of reducing carbon-intensive consumption below its preferred level. A larger ρ makes households less responsive to carbon prices, so that a given tax induces a smaller reduction in carbon use. We assume

Assumption 1.

$$\alpha_m > \alpha_e > \underline{x} > 0, \qquad \sigma_e > \sigma_m \geq 0.$$

Assumption 1 therefore implies that environmentalists have both a lower baseline demand for carbon-intensive consumption and a higher valuation of environmental quality than materialists.

The pre-tax price of the carbon-intensive good is normalized to one. A carbon tax $\tau_t \geq 0$ therefore raises its consumer price to $1 + \tau_t$. All adults receive the same exogenous income $w > 0$ and a uniform

[5] The quadratic specification is adopted for tractability and to make the change in behavioral incidence particularly transparent. The mechanism does not hinge on this particular functional form. Alternative smooth demand specifications, including logarithmic utility from carbon-intensive consumption, can generate the same policy-induced decline in behavioral distance and therefore the same cultural mechanism whenever behavioral distance crosses the transmission threshold.

lump-sum transfer R_t corresponding to recycled carbon-tax revenues.[6] Their budget constraint is

$$c_t^i + (1 + \tau_t)x_t^i = w + R_t. \tag{2}$$

We assume throughout that income is sufficiently high for the non-negativity constraint on numeraire consumption to remain slack over the range of policies considered below. Carbon-intensive consumption is subject to the lower bound

$$x_t^i \geq \underline{x}, \tag{3}$$

where $\underline{x} > 0$ denotes a lower bound on feasible carbon-intensive consumption.[7] We interpret $\underline{x}$ as a common short-run technological carbon frontier determined by the economic environment in which households operate. It captures the residual carbon-intensive consumption that households cannot avoid given the prevailing energy mix, transportation and housing infrastructure, supply chains, and the availability of low-carbon technologies. The frontier is common across cultural types because these constraints are properties of the shared technological and infrastructural environment rather than of households' environmental preferences. Lowering $\underline{x}$ therefore represents a decarbonization of that environment: for example, cleaner electricity, better public transportation, more energy-efficient housing, or broader access to low-carbon technologies can reduce the carbon use that remains unavoidable at a given date.[8] This frontier is distinct from the type-specific preference parameter α_i. The latter governs baseline demand for carbon-intensive consumption and, given our price normalization, coincides with optimal carbon use in the absence of taxation. By contrast, $\underline{x}$ restricts the feasible choice set. We take it to be common across cultural types in the benchmark, reflecting the technological and infrastructural environment they share rather than an intrinsic characteristic of either type. Since households are atomistic, they take the rebate R_t and environmental quality Q_t as given when choosing their own carbon-intensive consumption. Using (2) to substitute for c_t^i, the household's problem reduces to

$$\max_{x_t^i \geq \underline{x}} \left\{ -\tau_t x_t^i - \frac{\rho}{2}\left(x_t^i - \alpha_i\right)^2 \right\}. \tag{4}$$

In the absence of taxation, the household chooses its baseline level α_i. A positive carbon tax creates an incentive to move below this level, but doing so entails an increasing utility cost governed by ρ. The objective in (4) is strictly concave in x_t^i. The household's optimal carbon-intensive consumption is therefore

$$x_i^*(\tau_t) = \max\left\{ \underline{x}, \alpha_i - \frac{\tau_t}{\rho} \right\}. \tag{5}$$

Equation (5) highlights two margins of adjustment. When the household's choice is interior, a

[6] The recycling of carbon-pricing revenues through lump-sum transfers is standard in the literature; see, among others, Herweg and Schmidt (2022), Goulder et al. (2019), and Timilsina (2022). See also Besley and Persson (2023) for a model of carbon taxation and green values under electoral competition.

[7] The lower bound provides a transparent way of generating policy-induced behavioral compression but is not essential for the cultural mechanism. Section 3 shows that the reversal extends to general demand systems whenever equilibrium policy makes behavioral distance decline sufficiently to cross the transmission threshold.

[8] Using Danish register data, Gillingham and Munk-Nielsen (2019) show that households' driving responses to fuel prices depend strongly on the availability of substitution possibilities, in particular access to public transportation. Their results illustrate how transportation infrastructure conditions households' ability to reduce carbon-intensive consumption in response to higher prices.

higher carbon tax increases the cost of carbon-intensive consumption and reduces carbon use. Once the lower bound becomes binding, however, carbon-intensive consumption remains fixed at $\underline{x}$ and further increases in the tax no longer affect household behavior. The lower bound therefore determines the point at which the price instrument exhausts the household's adjustment margin. Since $\alpha_e < \alpha_m$, environmentalists have a lower baseline demand for carbon-intensive consumption and consume less than materialists whenever both types are unconstrained. Moreover, because the two types face the same lower bound and have the same responsiveness to the tax, environmentalists reach the constraint at a lower tax rate. For each cultural type, let $\bar{\tau}_i$ denote the carbon tax at which the lower bound becomes binding. Using (5), this threshold is

$$\bar{\tau}_i = \rho(\alpha_i - \underline{x}), \tag{6}$$

and since $\alpha_e < \alpha_m$, it follows that $\bar{\tau}_e < \bar{\tau}_m$. Indeed given that both cultural types face the same carbon tax, respond to it at the same rate $1/\rho$ and are subject to the same lower bound $\underline{x}$. Thus, the difference between the two thresholds does not come from a weaker response of environmentalists to the tax. It comes from the fact that, even in the absence of policy, environmentalists consume less of the carbon-intensive good than materialists. In other words, they have already undertaken part of the adjustment that the tax induces among materialists, such that their remaining adjustment margin, $\alpha_e - \underline{x}$, is smaller. Hence, the carbon tax exhausts the adjustment margin of environmentalists first.

Carbon-use regimes. Consequently, the economy moves through three regimes as the carbon tax increases: UU, where both types are unconstrained; BU, where environmentalists are constrained and materialists remain unconstrained; and BB, where both types are constrained. The UB regime cannot occur since any tax that makes the lower bound binding for materialists necessarily makes it binding for environmentalists as well. Let us note that the identity of those who adjust to the tax changes along this sequence. Indeed, below $\bar{\tau}_e$, both cultural types reduce their carbon-intensive consumption when the tax increases. However, once τ_t exceeds $\bar{\tau}_e$, environmentalists can no longer adjust, such that any further reduction in carbon use comes from materialists only. The following lemma summarizes household behavior in the three regimes.

Lemma 1. *Given a carbon tax $\tau_t \geq 0$, optimal carbon-intensive consumption is characterized as follows:*

1. *If $\tau_t < \bar{\tau}_e$, both types choose an interior solution and*

$$x_t^e = \alpha_e - \frac{\tau_t}{\rho}, \qquad x_t^m = \alpha_m - \frac{\tau_t}{\rho}. \tag{7}$$

   This is the UU regime.

2. *If $\bar{\tau}_e \leq \tau_t < \bar{\tau}_m$, the lower bound binds for environmentalists but not for materialists, so that*

$$x_t^e = \underline{x}, \qquad x_t^m = \alpha_m - \frac{\tau_t}{\rho}. \tag{8}$$

   This is the BU regime.

3. *If $\tau_t \geq \bar{\tau}_m$, the lower bound binds for both types and*

$$x_t^e = x_t^m = \underline{x}. \tag{9}$$

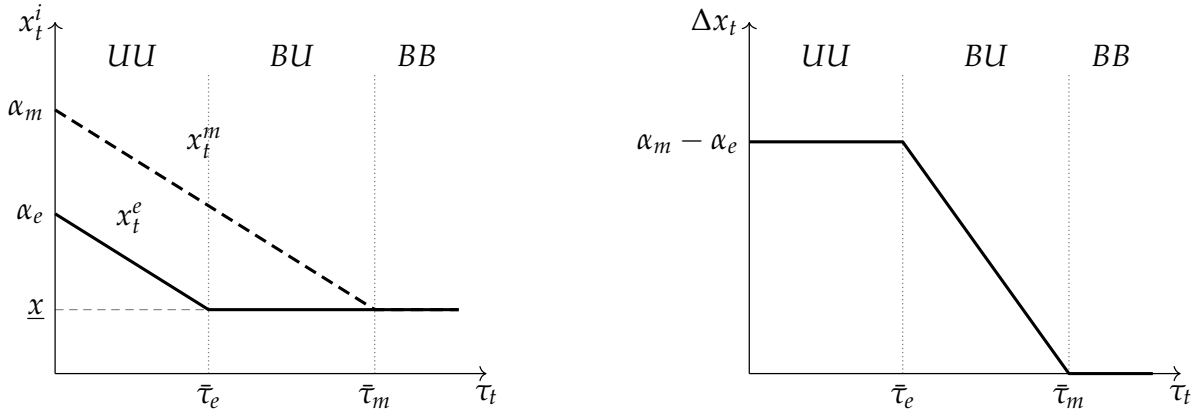
This is the BB regime.

Lemma 1 states that the way households respond to the carbon tax depends on the regime to which the economy belongs. In the UU regime, a higher tax reduces the carbon-intensive consumption of both cultural types by the same amount. In the intermediate BU regime, only materialists adjust. Finally, in the BB regime, both types consume $\underline{x}$, such that further increases in the tax no longer affect carbon-intensive consumption. Interestingly, this implies that the carbon tax affects not only the level of carbon-intensive consumption but also the extent to which the two cultural types behave differently. To capture this, let us define the behavioral gap as

$$\Delta x_t \equiv x_t^m - x_t^e. \tag{10}$$

The variable Δx_t accounts for the difference in carbon-intensive consumption between materialists and environmentalists. In other words, it measures how much environmental preferences translate into different behaviors: the higher Δx_t is, the more distinct the behaviors of environmentalists and materialists are. Conversely, when Δx_t decreases, the behaviors of the two cultural types converge. From there, using Lemma 1, the behavioral gap may be written as follows:

$$\Delta x(\tau_t) = \begin{cases} \alpha_m - \alpha_e, & \tau_t < \bar{\tau}_e, \\ \alpha_m - \underline{x} - \dfrac{\tau_t}{\rho}, & \bar{\tau}_e \leq \tau_t < \bar{\tau}_m, \\ 0, & \tau_t \geq \bar{\tau}_m. \end{cases} \tag{11}$$



(a) Carbon-intensive consumption, x_t^i.

(b) Behavioral gap, Δx_t.

Figure 1: Carbon pricing and behavioral convergence.

Figure 1 illustrates Lemma 1 and Equation (11). The behavioral gap remains equal to $\alpha_m - \alpha_e$ in the UU regime, decreases at rate $1/\rho$ in the BU regime and vanishes in the BB regime. Let us note that this convergence is not imposed by the policy instrument. Indeed, the carbon tax only changes the price faced by households, each household continuing to choose its optimal response to that price. In other words, the carbon tax does not eliminate cultural differences; it changes the incentives under which these differences are expressed in behavior.

So far, we have focused on the difference in behaviors between the two cultural types. Let us now

consider aggregate carbon use, $X_t = q_t x_t^e + (1 - q_t) x_t^m$, which determines environmental quality. Since the two cultural types do not adjust to the tax in the same way, the aggregate effect of the tax depends on the cultural composition of the population. Using Lemma 1, we obtain

$$X(\tau_t, q_t) = \begin{cases} q_t \alpha_e + (1 - q_t)\alpha_m - \dfrac{\tau_t}{\rho}, & \tau_t < \bar{\tau}_e, \\ q_t \underline{x} + (1 - q_t)\left(\alpha_m - \dfrac{\tau_t}{\rho}\right), & \bar{\tau}_e \leq \tau_t < \bar{\tau}_m, \\ \underline{x}, & \tau_t \geq \bar{\tau}_m. \end{cases} \tag{12}$$

Hence, the effect of the carbon tax on aggregate carbon use depends on the regime. In the UU regime, a marginal increase in the tax reduces aggregate carbon use by $1/\rho$, whatever the cultural composition of the population. In the BU regime, this effect is reduced to $(1 - q_t)/\rho$, since only the share of materialists still adjusts. In other words, the higher the share of environmentalists is, the weaker the aggregate response to the tax. Finally, in the BB regime, the carbon tax no longer affects aggregate carbon use. From there, we assume that environmental quality decreases with aggregate carbon use:

$$Q_t = \bar{Q} - \delta X_t, \qquad \delta > 0, \tag{13}$$

where $\bar{Q}$ denotes environmental quality in the absence of carbon use and δ measures the marginal damage associated with aggregate carbon-intensive consumption.[9] Hence, the carbon tax improves environmental quality as long as at least one cultural type still adjusts to it, that is, in the UU and BU regimes. Carbon-tax revenues are fully rebated to households through the lump-sum transfer R_t. Since there is a unit mass of households, the government budget constraint is given by

$$R_t = \tau_t X(\tau_t, q_t). \tag{14}$$

From there, substituting the optimal carbon-intensive consumption given by Lemma 1 into the utility function, we obtain the following indirect utility for a type-i household:

$$v_i(\tau_t, q_t) = w + R_t + \sigma_i Q_t + \begin{cases} -\alpha_i \tau_t + \dfrac{\tau_t^2}{2\rho}, & \tau_t < \bar{\tau}_i, \\ -\tau_t \underline{x} - \dfrac{\rho}{2}\left(\underline{x} - \alpha_i\right)^2, & \tau_t \geq \bar{\tau}_i. \end{cases} \tag{15}$$

Let us note that indirect utility is continuous at $\bar{\tau}_i$. Hence, the utility of a household varies continuously with the carbon tax, including when the economy moves from one regime to another. From there, it remains to determine the level of the carbon tax. Since this tax is chosen through the political process, it depends on the cultural composition of the electorate, such that, for a given q_t, it determines which of the regimes described in Lemma 1 prevails.

[9] We adopt a reduced-form flow specification of environmental damages in order to isolate the interaction between current policy, behavioral adjustment and cultural transmission. In particular, we abstract from the accumulation of atmospheric carbon and from the resulting intertemporal damage structure, which is central in quantitative climate-economy models (see, e.g., Golosov et al., 2014; Barrage, 2020). Such reduced-form environmental externalities are also standard in models studying the interaction between environmental behavior and endogenous preferences (see, for instance, Bezin, 2015).

2.2. The Political Economy of Carbon Pricing

We now turn to the political determination of the carbon tax. The cultural composition q_t is inherited from the previous generation, such that it is predetermined, and therefore taken as given, when the carbon tax is chosen. Following the probabilistic-voting approach, we assume that two office-motivated parties compete for votes by simultaneously proposing a carbon tax τ_t, and that idiosyncratic ideological shocks are identically distributed among environmentalists and materialists. Under this symmetry assumption, electoral competition weights each cultural group according to its population share.[10] Moreover, adults evaluate political platforms according to their own current-period utility and obtain no direct payoff from the cultural orientation of their child. In other words, although the current carbon tax affects the cultural composition of the next generation, this effect is not internalized when the tax is chosen. Electoral competition therefore leads both parties to maximize the average utility of the current adult generation, which, for a given q_t, is given by

$$\mathcal{W}(\tau_t, q_t) = q_t v_e(\tau_t, q_t) + (1 - q_t) v_m(\tau_t, q_t). \tag{16}$$

Since carbon-tax revenues are fully rebated, tax payments are a pure transfer within the current generation and do not constitute a net resource cost. The political trade-off is therefore between the environmental benefit of lower aggregate carbon use and the private cost borne by households when they reduce their carbon-intensive consumption below their preferred level. Using the household demands and the government budget constraint (14), the political objective can be written for each regime. When both cultural types are unconstrained, it is given by

$$\mathcal{W}^{UU}(\tau_t, q_t) = w + [q_t\sigma_e + (1 - q_t)\sigma_m]\left[\bar{Q} - \delta\left(q_t\alpha_e + (1 - q_t)\alpha_m\right)\right] + \frac{\delta}{\rho}\left[q_t\sigma_e + (1 - q_t)\sigma_m\right]\tau_t - \frac{\tau_t^2}{2\rho}. \tag{17}$$

When environmentalists are constrained while materialists are not, it becomes

$$\begin{aligned}\mathcal{W}^{BU}(\tau_t, q_t) = w &+ [q_t\sigma_e + (1 - q_t)\sigma_m]\left[\bar{Q} - \delta\left(q_t\underline{x} + (1 - q_t)\alpha_m\right)\right] - q_t\frac{\rho}{2}\left(\underline{x} - \alpha_e\right)^2 \\ &+ (1 - q_t)\left\{\frac{\delta}{\rho}\left[q_t\sigma_e + (1 - q_t)\sigma_m\right]\tau_t - \frac{\tau_t^2}{2\rho}\right\}.\end{aligned} \tag{18}$$

Finally, when both cultural types are constrained, it is given by

$$\mathcal{W}^{BB}(\tau_t, q_t) = w + [q_t\sigma_e + (1 - q_t)\sigma_m]\left(\bar{Q} - \delta\underline{x}\right) - q_t\frac{\rho}{2}\left(\underline{x} - \alpha_e\right)^2 - (1 - q_t)\frac{\rho}{2}\left(\underline{x} - \alpha_m\right)^2. \tag{19}$$

Note that $\mathcal{W}^{BB}$ does not depend on τ_t. Indeed, aggregate carbon use is fixed at $\underline{x}$, such that a higher tax neither improves environmental quality nor, since revenues are fully rebated, affects average utility. In other words, the political trade-off described above only exists in the UU and BU regimes, and we impose below a condition under which the BB regime is never reached. Maximizing $\mathcal{W}^{UU}$ and $\mathcal{W}^{BU}$ with respect to τ_t yields, for $q_t < 1$, the same preferred tax:

$$\tau^{UU}(q_t) = \tau^{BU}(q_t) = \delta\left[q_t\sigma_e + (1 - q_t)\sigma_m\right]. \tag{20}$$

[10] Indeed, with the same density of swing voters in the two groups, the common electoral responsiveness cancels out from the parties' first-order conditions; see Lindbeck and Weibull (1987), Persson and Tabellini (2000) and Besley and Persson (2023).

Indeed, in the BU regime, a marginal increase in the tax only affects materialists. Both the environmental gain and the private adjustment cost it generates are therefore scaled by the same factor $1 - q_t$, which leaves unchanged the tax rate at which they balance.

Hence, the two regimes in which households still adjust to the tax generate the same political rule. To determine which of them prevails for a given cultural composition, while ruling out the BB regime, we impose the following condition.

Assumption 2.

$$\delta\sigma_m < \rho(\alpha_e - \underline{x}) < \delta\sigma_e < \rho(\alpha_m - \underline{x}).$$

The first two inequalities ensure that both the UU and the BU regimes arise along the equilibrium path. Indeed, the preferred tax is equal to $\delta\sigma_m < \bar{\tau}_e$ when $q_t = 0$ and to $\delta\sigma_e > \bar{\tau}_e$ when $q_t = 1$, such that environmentalists exhaust their adjustment margin when the electorate is sufficiently green. The last inequality rules out the BB regime. Since the preferred tax never exceeds $\delta\sigma_e < \bar{\tau}_m$, materialists always retain an adjustment margin. This inequality also guarantees that the equilibrium tax is uniquely determined. Without it, whenever $\delta\left[q_t\sigma_e + (1 - q_t)\sigma_m\right] \geq \bar{\tau}_m$, the political objective would be maximized by any $\tau_t \geq \bar{\tau}_m$, since $\mathcal{W}^{BB}$ does not depend on τ_t. Then, we can claim that:

Lemma 2. *Suppose Assumption 2 holds. For every $q_t \in [0, 1)$, the unique symmetric political equilibrium is characterized by*

$$\tau_t^* = \delta\sigma_m + \delta(\sigma_e - \sigma_m)q_t. \tag{21}$$

Moreover, there exists a unique threshold

$$\bar{q} = \frac{\rho(\alpha_e - \underline{x}) - \delta\sigma_m}{\delta(\sigma_e - \sigma_m)} \in (0, 1), \tag{22}$$

such that the equilibrium is in the UU regime for $q_t < \bar{q}$ and in the BU regime for $\bar{q} \leq q_t < 1$.

Lemma 2 states that electoral competition implements the static Pigouvian tax evaluated at the average environmental valuation of the current electorate. Indeed, a marginal reduction in aggregate carbon use improves environmental quality by δ, an improvement valued on average at $q_t\sigma_e + (1 - q_t)\sigma_m$ by the current adult generation. This is the standard Pigouvian rule, abstracting from the dynamic and fiscal considerations emphasized in quantitative climate-economy models (Golosov et al., 2014; Barrage, 2020). Since $\sigma_e > \sigma_m$, the equilibrium tax increases with the environmentalist share: a greener electorate induces a more stringent carbon policy. Importantly, τ_t^* depends neither on the preferred levels of carbon-intensive consumption, α_e and α_m, nor on the lower bound $\underline{x}$. These parameters only determine which cultural type adjusts to the tax, and by how much.

The behavioral consequences of this political feedback, however, differ on both sides of $\bar{q}$. When $q_t < \bar{q}$, a larger environmentalist share raises the equilibrium tax and reduces the carbon use of both types at the same rate, such that the behavioral gap remains equal to $\alpha_m - \alpha_e$. Once q_t reaches $\bar{q}$, the equilibrium tax reaches $\bar{\tau}_e$ and environmentalists hit the lower bound. From there, further increases in q_t still raise the tax, but only materialists adjust, and their behavior progressively converges towards that of environmentalists. In other words, below $\bar{q}$, the political success of environmental values reduces carbon use without affecting the behavioral distinction between cultural types, whereas above $\bar{q}$, it progressively compresses this distinction. It is this change in the behavioral incidence of the same policy rule, rather than a change in the policy rule itself, that drives the cultural dynamics studied

below.

2.3. Payoffs and Behavioral Distinctiveness

Before turning to cultural transmission, we distinguish two consequences of the political equilibrium. A larger environmentalist share affects both the relative payoff associated with environmental preferences and the extent to which these preferences translate into distinct behaviors. As shown below, these two margins do not necessarily move in the same direction.

Let $V_i(q_t) \equiv v_i(\tau_t^*, q_t)$ denote the equilibrium payoff of a type-i adult, obtained by evaluating the indirect utility (15) at the equilibrium carbon tax, and let $D(q_t) \equiv V_e(q_t) - V_m(q_t)$ denote the relative payoff of environmentalists. Since the lump-sum transfer is common to all households, it cancels out from $D(q_t)$, which therefore reflects two channels.[11] First, as the environmentalist share increases, the equilibrium tax rises and aggregate carbon use falls. The resulting improvement in environmental quality benefits environmentalists relatively more, since $\sigma_e > \sigma_m$. Second, environmentalists consume less of the carbon-intensive good and are therefore less exposed to the increase in its tax-inclusive price. Both channels imply that $D(q_t)$ is strictly increasing in q_t, but their strength differs across regimes. In the UU regime, both cultural types adjust to the tax at the same rate, such that $D(q_t)$ is affine in q_t. In the BU regime, only materialists adjust. Hence, a higher tax improves environmental quality by less, and the advantage that environmentalists derive from their lower exposure to the tax, which is proportional to the behavioral gap, shrinks as materialists approach the lower bound. The relative payoff of environmentalists therefore keeps increasing, but at a decreasing rate.

The behavioral gap evolves differently. Combining (11) with Lemma 2, the behavioral gap along the political-equilibrium path is given by

$$\Delta x(q_t) = \begin{cases} \alpha_m - \alpha_e, & q_t < \bar{q}, \\ \alpha_m - \underline{x} - \dfrac{\delta\left[q_t\sigma_e + (1-q_t)\sigma_m\right]}{\rho}, & \bar{q} \leq q_t < 1. \end{cases} \tag{23}$$

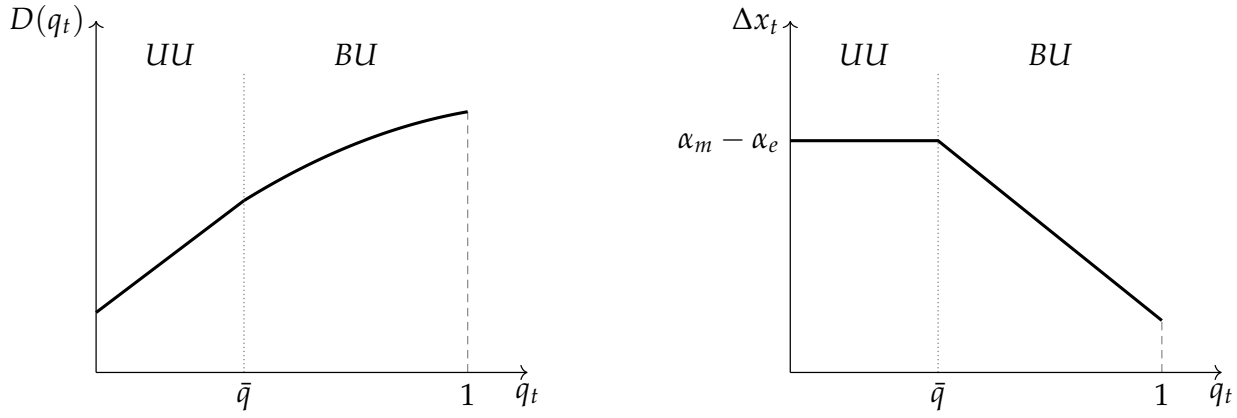
In other words, a larger environmentalist share leaves the behavioral gap unchanged as long as $q_t < \bar{q}$, but reduces it at rate $\delta(\sigma_e - \sigma_m)/\rho$ once $q_t \geq \bar{q}$, since the stronger carbon tax it induces only affects the behavior of materialists. From there, we can claim that:

Lemma 3. *Suppose Assumption 2 holds.*

1. *The relative equilibrium payoff of environmentalists, $D(q_t)$, is increasing in q_t.*
2. *The behavioral gap, $\Delta x(q_t)$, is constant for $q_t < \bar{q}$ and strictly decreasing for $\bar{q} \leq q_t < 1$.*

Lemma 3 is illustrated in Figure 2.

[11] The explicit expressions of $D(q_t)$ in the UU and BU regimes and the corresponding derivatives are reported in Appendix A.3.



(a) Relative payoff of environmentalists, $D(q_t)$.

(b) Behavioral gap, Δx_t.

Figure 2: Equilibrium payoffs and behavioral distinctiveness.

The contrast between the two panels is central to the cultural mechanism developed below. Once q_t exceeds $\bar{q}$, a larger environmentalist share keeps increasing the relative payoff of environmentalists, while the stronger carbon tax it induces makes materialists behave increasingly like them. In other words, above $\bar{q}$, economic incentives and behavioral distinctiveness move in opposite directions. Hence, if the transmission of environmental values weakens as q_t increases, it will not be because environmentalism becomes relatively less rewarding, but because the behavior associated with environmental preferences becomes less and less distinct from the behavior that climate policy induces among materialists.

3. Cultural Transmission

Here, we consider how environmental values are transmitted across generations and the associated cultural dynamics of the economy. We first describe the socialization process (Section 3.1) and then characterize the evolution of the cultural composition of the population (Section 3.2).

3.1. Socialization Process

We use the cultural transmission mechanism proposed by Besley (2017) (see also Besley and Ghatak, 2016; Besley and Persson, 2019a,b; Touré, 2021).[12] In our framework, a child acquires her cultural trait through the joint action of the family and society. A child is first socialized into her parent's cultural trait through the example of her parental role model. She is also socialized through the example of a role model randomly chosen in the adult population, the external role model, who is an environmentalist with probability q_t and a materialist with probability $1 - q_t$. If these two role models share the same trait, the child adopts it with probability one. Otherwise, the child has to choose between the two cultural traits.

Let us highlight that the main point of departure between our framework and that of Besley (2017) lies in the way this choice is made. In Besley (2017) and Besley and Persson (2019b), it is based on a noisy comparison between the utilities associated with each trait, such that cultural fitness is

[12] Baudin and Hiller (2019) propose an extension of this cultural transmission process to two populations. In Besley and Persson (2019b), who apply it to environmental values, the two cultural examples of a child are her two parents, and our protocol corresponds to their case without assortative matching.

payoff-based. In our framework, it is based on what the child can learn about the environmental trait from the behavior of her two role models, such that cultural fitness is *behavior-based*. This departure builds on three strands of the literature. First, a growing literature models cultural change as learning from the observed behavior of others (Fogli and Veldkamp, 2011; Fernández, 2013). Second, research on cultural evolution shows that values are transmitted more effectively by example than by words (Boyd and Richerson, 1985; Henrich, 2009).[13] Third, Bénabou and Tirole (2006) show that extrinsic incentives reduce the informational content of prosocial behavior about intrinsic motivation, a prediction supported experimentally by Ariely et al. (2009). We bring this insight into an intergenerational setting.

Closest to our approach, Bezin (2015) also makes the transmission of environmental preferences depend on observed behavior: children recognize the green trait of an adult only if this adult contributes to environmental quality, such that green agents who do not act appear brown. In her model, materialists never contribute, and green behavior becomes less informative when environmentalists do not act. We study the opposite margin. Environmentalists keep acting on their values, but carbon pricing induces materialists to behave like them, which makes environmental behavior less informative about environmental values. As in Carvalho (2013), the transmission of a trait depends on the behavioral difference it generates, such that a policy compressing behavioral differences across types can weaken the transmission of values. Unlike in his model, however, this policy is not exogenous: it is chosen by the current cultural composition of the electorate.

Information. The child knows the behavioral prescription of each trait: in the absence of constraints, environmentalists consume less of the carbon-intensive good than materialists by

$$\bar{r} \equiv \alpha_m - \alpha_e$$

at any given price. In other words, $\bar{r}$ is the behavioral gap between the two cultural types when neither of them is constrained. The child knows the declared trait of each of her two role models and observes, with noise, their carbon-intensive consumption, but not the constraints they face. Her signal is given by the behavioral difference between them,

$$\tilde{d} = \Delta x_t + \omega\xi, \qquad \xi \sim \mathcal{N}(0,1), \tag{24}$$

where $\omega > 0$ measures how imperfectly carbon-intensive behavior is observed. This creates an attribution problem. Since the child observes realized behavior but not the constraints shaping it, a behavioral gap $\Delta x_t < \bar{r}$ may be interpreted as evidence that the environmental trait has limited behavioral content, rather than as the consequence of a policy-induced behavioral convergence. This assumption is consistent with the evidence on the correspondence bias, according to which observers tend to attribute behavior to the dispositions of the actor rather than to the situational constraints she faces (Jones and Harris, 1967; Ross, 1977; Gilbert and Malone, 1995).

Trait choice. A child exposed to mixed influences (i.e., one whose parental and external role models have different traits) treats materialism as the default orientation: she adopts the environmental trait

[13] Children whose caregivers acted on their religious beliefs are more likely to hold these beliefs as adults, whereas religious teaching alone has little predictive power (Lanman and Buhrmester, 2017). In the environmental domain, advocates who adopt the behavior they promote are more persuasive (Kraft-Todd et al., 2018; Attari et al., 2016, 2019).

only if the behavior of her role models provides sufficient evidence that environmental values have observable behavioral content. Otherwise, she adopts the materialist trait.[14] This rule captures the logic of credibility-enhancing displays (Henrich, 2009): a value that its holders profess but do not act upon is not credible, and is therefore not adopted. Formally, the child compares the likelihood of her signal under two hypotheses: either the environmental trait shapes behavior as prescribed, such that $\tilde{d} \sim \mathcal{N}(\bar{r}, \omega^2)$, or it does not, such that $\tilde{d} \sim \mathcal{N}(0, \omega^2)$. She adopts the environmental trait if and only if the first hypothesis is at least as likely as the second, that is, if and only if

$$\tilde{d} \geq s \equiv \frac{\bar{r}}{2} = \frac{\alpha_m - \alpha_e}{2}. \tag{25}$$

The threshold s is the midpoint between the behavioral difference prescribed by environmental values and no difference at all. It depends on preferences only, not on policy.[15]

Hence, if we denote by Φ the cumulative distribution function of $\varepsilon \equiv \omega\xi \sim \mathcal{N}(0, \omega^2)$ and by ϕ the corresponding density function, the probability that a child exposed to mixed influences adopts the environmental trait equals $\Phi(\Delta x_t - s)$, with $\Phi(0) = 1/2$ and $\phi(\cdot) > 0$. Along the political-equilibrium path, the behavioral gap is given by (23), such that $\Delta x_t = \Delta x(q_t)$. We can then claim that:

Lemma 4. *Under mixed influences, the probability that the child adopts the environmental trait is given by* $\Phi(\Delta x(q_t) - s)$. *Moreover,*

$$\Phi(\Delta x(q_t) - s) \gtrless \frac{1}{2} \quad \Longleftrightarrow \quad \Delta x(q_t) \gtrless s.$$

Lemma 4 has the same reduced-form structure as the transmission rule in Besley (2017) and Besley and Persson (2019b) (see also Touré, 2021; Baudin and Hiller, 2019), with the behavior-based fitness advantage $\Delta x(q_t) - s$ replacing the payoff difference. In other words, the environmental trait has a positive fitness advantage if and only if environmentalists and materialists differ in their behavior by more than half the difference prescribed by environmental values. This makes the cultural effect of carbon pricing transparent. As long as both types are unconstrained, $\Delta x(q_t) = \bar{r} = 2s$, such that the environmental trait has a positive fitness advantage. Once environmentalists reach the lower bound, however, carbon pricing reduces the behavioral gap by moving materialists towards them.[16] Since children observe behavior but not the constraints generating it, this convergence makes environmental values appear less behaviorally consequential. When the tax has closed half of the initial behavioral gap, the environmental trait loses its fitness advantage, even though the preferences of environmentalists are unchanged. Materialist compliance thus makes environmentalist conviction harder to observe.[17]

[14] Materialism plays the role of the default trait, acquired in the absence of successful socialization to the alternative, as in models of green socialization (Bezin, 2015, 2019), in Carvalho (2013), and in our extension with parental socialization (Section 6).

[15] If the child attaches prior probability $h \in (0, 1)$ to the hypothesis that the environmental trait does not shape behavior, and adopts it when its posterior probability is at least one half, the threshold becomes $s = \bar{r}/2 + (\omega^2/\bar{r}) \ln[h/(1-h)]$. Greater skepticism raises s and, by Proposition 1, lowers the long-run share of environmentalists. All results hold for any $s \in (\alpha_m - \underline{x} - \delta\sigma_e/\rho,\ \alpha_m - \alpha_e)$.

[16] The lower bound is not the source of the reversal. Since, by Lemma 4, the fitness advantage of the environmental trait only depends on the sign of $\Delta x(q_t) - s$, the same reversal arises whenever the equilibrium carbon tax compresses the behavioral gap below s, provided that children do not attribute this compression to policy. The lower bound only provides a transparent way of generating such a compression.

[17] The reversal relies on the fact that children do not observe the constraints faced by their role models. Suppose instead that the child observes these constraints and expects the equilibrium gap $\Delta x(q_t)$, rather than $\bar{r}$, under the first hypothesis. Behavioral convergence still weakens transmission, since a smaller gap is harder to distinguish from no difference. However, a reversal then additionally requires prior skepticism. If the child attaches prior probability $h > 1/2$ to the hypothesis that the environmental trait does not shape behavior, the adoption probability becomes $N\big((\Delta x^2 - \bar{s}^2)/(2\omega\Delta x)\big)$, where N

3.2. Cultural Evolution

Using this framework and denoting by p_t^{ij} the probability that a child whose parent has trait $i \in \{e, m\}$ adopts trait $j \in \{e, m\}$ on date t, we obtain the following transition probabilities:

$$\begin{aligned} p_t^{ee} &= q_t + (1 - q_t)\Phi(\Delta x(q_t) - s), \quad & p_t^{em} &= (1 - q_t)\left[1 - \Phi(\Delta x(q_t) - s)\right], \\ p_t^{me} &= q_t\Phi(\Delta x(q_t) - s), & p_t^{mm} &= (1 - q_t) + q_t\left[1 - \Phi(\Delta x(q_t) - s)\right]. \end{aligned} \tag{26}$$

The result is the following equation of evolution of q_t:

$$q_{t+1} = q_t p_t^{ee} + (1 - q_t)p_t^{me} = q_t^2 + 2q_t(1 - q_t)\Phi(\Delta x(q_t) - s) \equiv G(q_t). \tag{27}$$

The first term on the right side, q_t^2, corresponds to the probability that a child has parental and external role models who are both environmentalists. In this case, the child adopts the environmental trait with probability one. The second term equals the probability that a child has role models with different traits, $2q_t(1 - q_t)$, multiplied by the probability of adopting the environmental trait in this case, $\Phi(\Delta x(q_t) - s)$. Rearranging (27), we can derive the law of motion of q_t as follows:

$$q_{t+1} - q_t = q_t(1 - q_t)\left\{2\Phi(\Delta x(q_t) - s) - 1\right\}. \tag{28}$$

This equation corresponds to replicator dynamics, which are standard in evolutionary models of cultural change (see, e.g., Besley, 2017; Bisin and Verdier, 2011). According to (28), the evolution of q_t depends on the behavioral gap:

$$q_{t+1} \gtrless q_t \quad \Longleftrightarrow \quad \Delta x(q_t) \gtrless s. \tag{29}$$

As we know from our analysis of the political equilibrium, $\Delta x(q_t)$ crucially depends on the value of q_t in comparison to the threshold $\bar{q}$ (Lemma 3). First, let us consider the case of $q_t < \bar{q}$. Both cultural types are unconstrained, such that $\Delta x(q_t) = \bar{r} = 2s$, and environmental values spread. Second, if $q_t \geq \bar{q}$, environmentalists are constrained, and the behavioral gap decreases at rate $\delta(\sigma_e - \sigma_m)/\rho$, until it reaches $\alpha_m - \underline{x} - \delta\sigma_e/\rho$ as q_t approaches one. Whether the fitness advantage of the environmental trait eventually becomes negative therefore depends on how strongly a fully environmentalist electorate would tax carbon. We impose the following condition.

Assumption 3.

$$\delta\sigma_e > \rho\left(\frac{\alpha_m + \alpha_e}{2} - \underline{x}\right).$$

The left-hand side of this condition is the carbon tax chosen by a fully environmentalist electorate, $\tau^*(1) = \delta\sigma_e$ (Lemma 2). The right-hand side is the carbon tax that reduces the behavioral gap to half of its initial value. Indeed, in the *BU* regime, the behavioral gap is equal to $\alpha_m - \underline{x} - \tau_t/\rho$, such that it reaches the learning threshold $s = (\alpha_m - \alpha_e)/2$ at the carbon tax

$$\tau^C = \rho\left(\frac{\alpha_m + \alpha_e}{2} - \underline{x}\right) = \frac{\bar{\tau}_e + \bar{\tau}_m}{2}. \tag{30}$$

denotes the standard normal distribution and $\tilde{s} \equiv \omega[2\ln(h/(1-h))]^{1/2}$. This probability increases with Δx and is below one half if and only if $\Delta x < \tilde{s}$. Hence, the weakening of transmission is robust to the sophistication of children, whereas its reversal requires either an attribution problem or a skeptical prior. In the latter case, the direction of cultural change is determined by the sign of $\Delta x(q_t) - \tilde{s}$, such that the analysis below applies with $\tilde{s}$ in place of s.

Assumption 3 can therefore be written as $\delta\sigma_e > \tau^C$: a fully environmentalist electorate would choose a carbon tax high enough to close more than half of the initial behavioral gap. Without this condition, the behavioral gap would remain above s for all $q_t < 1$, and environmental values would spread to the whole population. Behavioral convergence would then slow down this diffusion beyond $\bar{q}$, but would not reverse it. Note also that τ^C lies exactly halfway between $\bar{\tau}_e$ and $\bar{\tau}_m$, such that Assumption 3 is compatible with Assumption 2. Combining the two assumptions, we obtain $\bar{\tau}_e < \tau^C < \delta\sigma_e < \bar{\tau}_m$. In other words, the carbon tax closes half of the initial behavioral gap after environmentalists have reached the lower bound, but before materialists do. The properties of cultural dynamics are summarized in the following proposition.

Proposition 1. *Suppose Assumptions 2 and 3 hold.*

(i) There exists a unique interior steady state $q_H^* \in (\bar{q}, 1)$, characterized by $\Delta x(q_H^*) = s$, or equivalently $\tau^*(q_H^*) = \tau^C$. It is given by

$$q_H^* = \frac{\tau^C - \delta\sigma_m}{\delta(\sigma_e - \sigma_m)} = \bar{q} + \frac{\rho(\alpha_m - \alpha_e)}{2\delta(\sigma_e - \sigma_m)}. \tag{31}$$

(ii) The environmentalist share increases for $0 < q_t < q_H^*$ and decreases for $q_H^* < q_t < 1$.

(iii) The steady state q_H^* is locally stable if and only if

$$q_H^*(1 - q_H^*)\,\phi(0)\,\frac{\delta(\sigma_e - \sigma_m)}{\rho} < 1, \qquad \text{with } \phi(0) = \frac{1}{\sqrt{2\pi}\,\omega}. \tag{32}$$

Proposition 1 states that the environmental trait spreads as long as the carbon tax has closed less than half of the initial behavioral gap, and declines beyond. Indeed, for $q_t < q_H^*$, we have $\Delta x(q_t) > s$, such that the environmental trait has a positive fitness advantage, although this advantage declines once q_t exceeds $\bar{q}$. At q_H^*, the two hypotheses are equally likely given the average signal, and $\Phi(\Delta x(q_H^*) - s) = 1/2$. Beyond q_H^*, the fitness advantage becomes negative and the environmentalist share declines. Part (iii) follows from

$$G'(q_H^*) = 1 - 2q_H^*(1 - q_H^*)\,\phi(0)\,\frac{\delta(\sigma_e - \sigma_m)}{\rho},$$

which is always below one, and exceeds -1 if and only if (32) holds. When carbon-intensive behavior is highly observable (ω small), cultural fitness reacts strongly to small changes in the behavioral gap. In this case, $G'(q_H^*)$ may become negative, such that the environmentalist share and the carbon tax oscillate around q_H^*, and the steady state may even become unstable if (32) is violated.

Importantly, this reversal is not driven by a deterioration in the relative economic position of environmentalists. As shown in Lemma 3, their relative payoff keeps increasing in the *BU* regime, such that $D'(q_t) > 0$ while $\Delta x'(q_t) < 0$. Under a payoff-based cultural fitness, as in Besley (2017) and Besley and Persson (2019b), the fitness advantage of environmentalism would therefore keep increasing with q_t, such that a larger environmentalist share could not turn it negative. Under a behavior-based cultural fitness, it can. In other words, climate policy can improve the relative economic position of environmentalists while making their values harder to learn from their behavior.

These properties are also illustrated in Figure 3, which depicts the mapping $G(q_t)$ together with the mapping that would prevail if the behavioral gap were fixed to $\bar{r}$, that is, in the absence of

behavioral convergence. The two mappings coincide up to $\bar{q}$. Beyond $\bar{q}$, the carbon tax moves materialists towards environmentalists, the signal received by children weakens, and $G(q_t)$ falls below the mapping obtained with a fixed behavioral gap. In the absence of behavioral convergence, the environmental trait would keep its fitness advantage and spread to the whole population. With behavioral convergence, $G(q_t)$ crosses the 45° line at q_H^*.

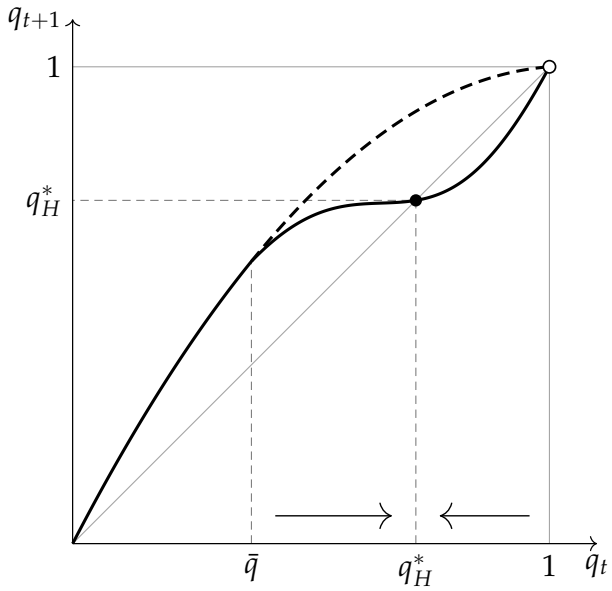


Figure 3: Dynamics of environmental values. The solid curve depicts the mapping $G(q_t)$. The dashed curve depicts the mapping obtained when the behavioral gap is fixed to $\bar{r}$.

The figure also highlights the different roles of $\bar{q}$ and q_H^*. The first threshold marks the change in the behavioral incidence of carbon pricing: beyond it, materialists alone supply additional abatement, and the fitness advantage of the environmental trait starts to decline. The second threshold marks the point at which this advantage becomes negative. In between, the environmentalist share keeps rising while the forces sustaining its diffusion are already weakening.

This mechanism complements the coevolutionary perspective of Besley and Persson (2019b) and Besley and Persson (2023) (see also Bezin, 2019), in which environmental values and the economic environment can reinforce one another during a green transition. In our model, by contrast, the same political feedback may become self-limiting. A larger environmentalist population keeps supporting stronger carbon pricing, but once environmentalists are constrained, the resulting policy operates through the behavior of materialists and reduces the behavior-based fitness of environmental values. Relative to Carvalho (2013), the policy that compresses behavioral differences is itself generated by the cultural composition of the electorate, and the strength of this compression changes endogenously once environmentalists cease to be marginal adjusters.

Finally, cultural change feeds back into future climate policy. Since $\tau_t^* = \delta\left[q_t\sigma_e + (1-q_t)\sigma_m\right]$, we have

$$\tau_{t+1}^* - \tau_t^* = \delta(\sigma_e - \sigma_m)\left(q_{t+1} - q_t\right). \tag{33}$$

When environmental values spread, the future constituency supporting carbon pricing expands. Conversely, when behavioral convergence leads the environmentalist share to decline, the next generation selects a lower carbon tax. Note that neither link is designed to shape culture: the cultural

effect of climate policy is an unintended consequence of a contemporaneous electoral equilibrium. Hence, current abatement and the future constituency for climate policy may move in opposite directions.

4. Policy Design

Section 3 has shown that carbon pricing can reverse the transmission of environmental values by compressing the behavioral difference between environmentalists and materialists. In this section, we study the implications of this mechanism for the design of climate policy. Three results emerge. First, the onset of the cultural reversal can be characterized in policy space by a threshold level of the carbon tax, which is crossed while the tax still reduces current carbon use. Second, policies generating the same current reduction in carbon use can have opposite effects on the future distribution of environmental values. Third, expanding the low-carbon adjustment possibilities of households has different cultural consequences from compensating them for the financial burden of carbon pricing. Taken together, these results separate the current effectiveness of climate policy from its consequences for the intergenerational transmission of environmental values.

4.1. The Carbon-Tax Threshold

Proposition 1 characterizes the cultural reversal in terms of the environmentalist share. Since the equilibrium tax is strictly increasing in q_t, the same reversal can be expressed in terms of the carbon tax. Indeed, from Proposition 1, $\tau^*(q_H^*) = \tau^C$, where τ^C is given by (30). Along the political-equilibrium path, environmental values therefore spread when $\tau_t^* < \tau^C$ and decline when $\tau_t^* > \tau^C$. In other words, τ^C is the carbon tax at which policy-induced compliance has compressed the behavioral gap exactly to the level at which neither cultural trait enjoys a fitness advantage under mixed influences.

Let us note that τ^C is neither the tax at which behavioral convergence begins nor the tax at which carbon pricing ceases to affect behavior. From Assumptions 2 and 3, we have $\bar{\tau}_e < \tau^C < \bar{\tau}_m$, which has two implications. First, there is a non-degenerate range of carbon taxes, $(\bar{\tau}_e, \tau^C)$, in which policy already compresses behavioral differences while environmental values keep spreading. The onset of behavioral convergence and the onset of cultural decline are therefore distinct events: the cultural effect of carbon pricing deteriorates gradually before the direction of cultural change reverses. Second, materialists still retain an adjustment margin when τ^C is crossed. Above τ^C, a higher tax thus keeps reducing carbon use and improving current environmental quality. In other words, the cultural reversal does not arise because the price instrument has become ineffective in the usual static sense, but while it continues to achieve its contemporaneous objective.

Importantly, the existence of τ^C should not be interpreted as implying that sufficiently strong carbon pricing should be avoided. What matters is the behavioral margin through which additional abatement is obtained. Once environmentalists are constrained, a higher carbon tax reduces carbon use by moving materialists towards environmentalist behavior. Other interventions may obtain the same abatement by expanding the adjustment possibilities of environmentalists instead. We now turn to this comparison.

4.2. Behavioral Incidence: Same Abatement, Opposite Cultural Effects

The previous result shows that the cultural effect of climate policy cannot be inferred from its effect on current carbon use alone. To isolate this point, we compare two marginal interventions that generate the same reduction in aggregate carbon use but induce adjustment by different cultural groups.

Consider an economy strictly inside the *BU* regime and hold the cultural composition q fixed. Environmentalists are constrained at $\underline{x}$, whereas materialists choose $\alpha_m - \tau/\rho$. Aggregate carbon use and the behavioral gap are therefore given by

$$X = q\underline{x} + (1-q)\left(\alpha_m - \frac{\tau}{\rho}\right), \qquad \Delta x = \alpha_m - \underline{x} - \frac{\tau}{\rho}. \tag{34}$$

The two interventions operate on opposite sides of the behavioral gap. A higher carbon tax leaves environmentalists at $\underline{x}$ and induces materialists to reduce their carbon use, such that Δx falls. By contrast, a lower $\underline{x}$ relaxes the constraint faced by environmentalists and allows them to reduce their carbon use further, while leaving the choice of materialists unchanged, such that Δx rises.

Let $\eta > 0$ denote a marginal reduction in aggregate carbon use. From (34), obtaining this reduction through the carbon tax requires $d\tau = \rho\eta/(1-q)$, which reduces the behavioral gap by $\eta/(1-q)$. Obtaining it through the lower bound requires $d\underline{x} = -\eta/q$, which increases the behavioral gap by η/q. Since, holding q fixed, the cultural mapping (27) responds to the behavioral gap with slope $2q(1-q)\phi(\Delta x - s)$, we can claim that:

Proposition 2. *Consider an economy strictly inside the BU regime and two marginal interventions that generate the same reduction $\eta > 0$ in current aggregate carbon use. The tax-induced reduction and the reduction obtained by lowering $\underline{x}$ change the next-generation environmentalist share by, respectively,*

$$dq^{\tau}_{t+1} = -2q\,\phi(\Delta x - s)\,\eta < 0 \qquad \textit{and} \qquad dq^{\underline{x}}_{t+1} = 2(1-q)\,\phi(\Delta x - s)\,\eta > 0. \tag{35}$$

Proposition 2 states that equal current abatement can have opposite cultural effects. By construction, the two interventions generate the same reduction in current carbon use, and therefore the same improvement in current environmental quality, and neither of them changes preferences. Their cultural effects differ because the marginal adjustment is supplied by different cultural groups. When the additional abatement is obtained through a stronger price signal, materialists supply it and become less distinct from environmentalists, which weakens the transmission of environmental values. When the same abatement is obtained by expanding low-carbon adjustment possibilities, environmentalists supply it and become more distinct from materialists, which strengthens this transmission. In other words, aggregate abatement is not a sufficient statistic for the cultural effect of climate policy: what matters is its behavioral incidence.

The magnitudes in Proposition 2 also have a natural interpretation. The term $\phi(\Delta x - s)$ measures the sensitivity of cultural transmission to a marginal change in the behavioral gap at the prevailing cultural state. The factors q and $1-q$ come from the composition of the population. For a given aggregate abatement, adjustment by a small group requires a larger change in the behavior of each of its members. This individual response offsets the weight of the group in aggregate carbon use, such that the cultural effect of tax-induced abatement is proportional to q, and that of the lower bound to $1-q$. Note that the sign of both effects does not depend on the distribution underlying Φ.

This result introduces an intergenerational dimension into the comparison of climate-policy instruments. Two interventions may be environmentally equivalent from the perspective of the current generation while generating different cultural compositions in the next one. Since the equilibrium tax increases with the environmentalist share, this difference also affects the political support for future climate policy, as shown by (33). Hence, identical abatement today can place the economy on different future policy paths. This provides a distinct rationale for combining carbon pricing with policies that expand low-carbon adjustment possibilities. Existing work emphasizes complementarities between carbon prices, directed technological change and the availability of cleaner substitutes (Acemoglu et al., 2012; Aghion et al., 2016), while Mattauch et al. (2022) show that endogenous preferences can alter the relative desirability of climate-policy instruments. The mechanism here is different: even holding preferences and current abatement fixed, the composition of the policy mix matters because different instruments assign the marginal behavioral adjustment to different cultural groups.

Let us note that Proposition 2 is not a welfare ranking of the two instruments. The comparison holds current abatement fixed and does not model the resource cost of reducing $\underline{x}$. Its purpose is narrower: it identifies a dynamic cultural effect that remains invisible in a static comparison based on aggregate emissions.

4.3. Carbon Dependence, Adaptation, and Compensation

The previous result suggests that the cultural consequences of climate policy depend not only on the strength of carbon pricing, but also on the extent to which households can substitute away from carbon-intensive consumption. As shown by (5), the lower bound $\underline{x}$ captures carbon dependence: it determines the point at which the price instrument exhausts the adjustment margin of a household. Policies that reduce $\underline{x}$ can therefore be interpreted as expanding the set of low-carbon alternatives available to households. In practice, public transportation, charging infrastructure, energy-efficient housing, cleaner electricity and low-carbon heating technologies all reduce the carbon-intensive consumption that remains unavoidable in the prevailing economic environment. Evidence that transportation infrastructure shapes the response of households to fuel prices is consistent with this interpretation (Gillingham and Munk-Nielsen, 2019).

Reducing carbon dependence has three effects. First, from (22) and (31), a lower $\underline{x}$ raises both $\bar{q}$ and q_H^*. In other words, it postpones the point at which environmentalists exhaust their adjustment margin and allows environmental values to keep spreading over a larger range of cultural compositions. If $\underline{x}$ falls enough for Assumption 3 to be violated, the cultural reversal disappears altogether.[18] Second, in the BU regime, a lower $\underline{x}$ directly reduces the carbon use of environmentalists and therefore current aggregate carbon use, as shown by (12). Third, it increases the behavioral gap (11), which strengthens the transmission of environmental values through (27). Through the political feedback (33), the resulting larger environmentalist share then supports a more stringent carbon tax in the future. Hence, expanding low-carbon alternatives generates a long-run environmental effect beyond its immediate impact on current carbon use. Importantly, these policies do not affect preferences: their cultural effect operates by changing the extent to which existing preferences can be expressed through behavior.

This mechanism also clarifies why compensation and policies that shift the lower bound are not equivalent. Compensation changes the resources of households without shifting the set of feasible

[18] Similarly, if $\underline{x}$ falls enough for $\rho(\alpha_e - \underline{x}) > \delta\sigma_e$, environmentalists are never constrained along the equilibrium path and behavioral convergence does not occur.

low-carbon behaviors. Since the household choice (5) does not depend on lump-sum transfers, compensation leaves carbon demand, the behavioral gap, and the thresholds $\bar{q}$ and q_H^* unchanged. By contrast, a lower $\underline{x}$ expands the set of feasible low-carbon behaviors itself and, in the BU regime, restores an adjustment margin for environmentalists. This does not imply that compensation is irrelevant for climate policy. Revenue recycling and the perceived incidence of carbon taxation matter for distributional reasons and for political acceptability (Klenert et al., 2018; Douenne and Fabre, 2022). The distinction here concerns a different margin: transfers can offset the financial burden of a higher carbon price, but they do not create public transportation, charging infrastructure or other low-carbon substitutes. Let us also note that, unlike in Mattauch et al. (2022), infrastructure does not change preferences in our model. Low-carbon alternatives affect cultural dynamics because they change how given preferences translate into behavior.

To sum up, carbon pricing induces adjustment by materialists in the BU regime, compensation changes resources without affecting behavioral differences, and low-carbon alternatives restore an adjustment margin for constrained environmentalists. Together with Proposition 2, this shows that the composition of the climate-policy mix matters even when its current environmental effect is held fixed, since policies generating the same reduction in carbon use today differ in whose behavior adjusts. The next section turns to the normative implications of this mechanism, and asks whether the political tax remains efficient once its effect on the cultural composition of future generations is taken into account.

5. Welfare Analysis

So far, our analysis has been positive. Climate policy is chosen by the current electorate, and the cultural composition of the population evolves as an unintended consequence of the behavioral response to this policy. In this section, we study whether this feedback generates an intergenerational externality. This question is particularly relevant in our framework since the political equilibrium is not statically distorted. As shown in Lemma 2, the equilibrium carbon tax,

$$\tau_t^* = \tau^*(q_t) = \delta\left[q_t\sigma_e + (1-q_t)\sigma_m\right],$$

is the Pigouvian tax evaluated at the average environmental valuation of the current electorate: for a given cultural composition q_t, it maximizes the average utility of the current generation. The inefficiency studied below has a different source. The current carbon tax also affects the cultural composition of the next generation, q_{t+1}, and therefore future carbon demand and future climate policy. This effect is not taken into account by current voters.

The section is organized as follows. Section 5.1 defines the intertemporal welfare criterion and characterizes welfare along the political rule. Section 5.2 shows that, once environmentalists are constrained, the carbon tax has a cultural cost that voters do not take into account, such that the political tax is too high. Section 5.3 compares complete policy paths, and Section 5.4 shows that this comparison does not depend on the endogeneity of preferences.

5.1. Intertemporal Welfare

Current welfare. Let $\mathcal{W}(\tau_t, q_t) = q_t v_e(\tau_t, q_t) + (1-q_t)v_m(\tau_t, q_t)$ denote the average utility of the generation of adults living at date t, as in (16). Since carbon-tax revenues are fully rebated, tax payments cancel out with the lump-sum transfer when utilities are aggregated. Using (14), current

welfare may therefore be written as follows:

$$\mathcal{W}(\tau_t, q_t) = w + [q_t\sigma_e + (1-q_t)\sigma_m]\, Q(\tau_t, q_t) - q_t\frac{\rho}{2}\left(x_t^e - \alpha_e\right)^2 - (1-q_t)\frac{\rho}{2}\left(x_t^m - \alpha_m\right)^2, \tag{36}$$

where $Q(\tau_t, q_t) = \bar{Q} - \delta X(\tau_t, q_t)$ is environmental quality and x_t^i is the carbon-intensive consumption of type-i households given by Lemma 1. The second term is the value of environmental quality, weighted by the average environmental valuation of the current generation. The last two terms are the adjustment costs borne by environmentalists and materialists, that is, the utility losses they incur when they consume less of the carbon-intensive good than their preferred level.

Let us make (36) explicit in the *BU* regime, which is the relevant one for our analysis. In this regime, environmentalists consume $x_t^e = \underline{x}$, which costs them $\rho(\alpha_e - \underline{x})^2/2 = \bar{\tau}_e^2/(2\rho)$, while materialists consume $x_t^m = \alpha_m - \tau_t/\rho$, which costs them $\tau_t^2/(2\rho)$. Aggregate carbon use is $X(\tau_t, q_t) = \underline{x} + (1 - q_t)\Delta x(\tau_t)$, where $\Delta x(\tau_t) = \alpha_m - \underline{x} - \tau_t/\rho$ is the behavioral gap. Current welfare is then given by

$$\mathcal{W}(\tau_t, q_t) = w + [q_t\sigma_e + (1-q_t)\sigma_m]\left[\bar{Q} - \delta\underline{x} - \delta(1-q_t)\Delta x(\tau_t)\right] - q_t\frac{\bar{\tau}_e^2}{2\rho} - (1-q_t)\frac{\tau_t^2}{2\rho}. \tag{37}$$

Cultural transition. To evaluate policies that depart from the political rule, we need to know how any carbon tax τ_t affects the cultural composition of the next generation. From Section 3, the share of environmentalists at date $t+1$ is given by

$$q_{t+1} = H(q_t; \tau_t) \equiv q_t^2 + 2q_t(1-q_t)\Phi(\Delta x(\tau_t) - s), \tag{38}$$

where the behavioral gap $\Delta x(\tau_t)$ is given by (11). The function H describes how the cultural composition evolves for any carbon tax, and not only for the tax chosen by voters. Along the political-equilibrium path, $\tau_t = \tau^*(q_t)$, so that the cultural law of motion (27) is the special case $G(q_t) = H(q_t; \tau^*(q_t))$.

Intertemporal welfare. A policy path π is a sequence of carbon taxes $\{\tau_t^\pi\}_{t\geq 0}$. Starting from q_0, it generates a sequence of cultural compositions $\{q_t^\pi\}_{t\geq 0}$, with $q_{t+1}^\pi = H(q_t^\pi; \tau_t^\pi)$. The discounted welfare associated with this path is

$$\mathcal{V}^\pi(q_0) = \sum_{t=0}^{\infty} \beta^t \mathcal{W}(\tau_t^\pi, q_t^\pi), \qquad \beta \in (0,1). \tag{39}$$

The discount factor β is a social evaluation parameter. It does not represent the altruism of parents and does not affect the political equilibrium, since voters keep evaluating platforms according to their current utility only. The cultural composition does not enter this criterion directly; it only matters through the utilities, carbon use, environmental quality and adjustment costs that it generates in the future.

Welfare along the political rule. When the tax is set according to the political rule, current welfare and environmental quality only depend on the cultural composition. We denote them by $\mathcal{W}^P(q_t) \equiv$

$\mathcal{W}(\tau^*(q_t), q_t)$ and $Q^P(q_t) \equiv Q(\tau^*(q_t), q_t)$. From (37), we obtain, in the *BU* regime,

$$\mathcal{W}^P(q_t) = w + \frac{\tau^*(q_t)}{\delta} Q^P(q_t) - q_t \frac{\bar{\tau}_e^2}{2\rho} - (1-q_t)\frac{\tau^*(q_t)^2}{2\rho}, \qquad Q^P(q_t) = \bar{Q} - \delta \underline{x} - \delta(1-q_t)\Delta x(q_t), \quad (40)$$

where we use $q_t\sigma_e + (1-q_t)\sigma_m = \tau^*(q_t)/\delta$ and $\Delta x(q_t) = \alpha_m - \underline{x} - \tau^*(q_t)/\rho$. The welfare of an economy that starts from q_t and follows the political rule forever is then

$$J^P(q_t) = \sum_{k=0}^{\infty} \beta^k \mathcal{W}^P(q_{t+k}), \qquad \text{with } q_{t+k+1} = G(q_{t+k}). \quad (41)$$

This continuation welfare satisfies the recursive relation $J^P(q_t) = \mathcal{W}^P(q_t) + \beta J^P(q_{t+1})$, with $q_{t+1} = G(q_t)$. The following lemma characterizes how current welfare varies with the cultural composition along the political rule.

Lemma 5 (Welfare along the political rule)**.** *In the BU regime,*

$$\frac{d\mathcal{W}^P(q_t)}{dq_t} = (\sigma_e - \sigma_m)Q^P(q_t) + \tau^*(q_t)\Delta x(q_t) + \frac{\tau^*(q_t)^2 - \bar{\tau}_e^2}{2\rho}. \quad (42)$$

Consequently, $\mathcal{W}^P(q_t)$ is strictly increasing in q_t in any part of the BU regime in which $Q^P(q_t) \geq 0$.

Proof. See Appendix D.1. □

The derivation of (42) is simple. Since $\tau^*(q_t)$ maximizes $\mathcal{W}(\tau_t, q_t)$, the effect of q_t that goes through the tax is equal to zero (envelope theorem), and $d\mathcal{W}^P(q_t)/dq_t$ is the partial derivative of (37) with respect to q_t, evaluated at $\tau_t = \tau^*(q_t)$. This derivative gives the welfare effect of replacing one materialist by one environmentalist in the current generation, and it is made of three terms. First, an environmentalist attaches a higher value to environmental quality than a materialist, which gives $(\sigma_e - \sigma_m)Q^P(q_t)$. Second, an environmentalist consumes $\Delta x(q_t)$ less of the carbon-intensive good than a materialist. This improves environmental quality by $\delta\Delta x(q_t)$, an improvement valued at $q_t\sigma_e + (1-q_t)\sigma_m = \tau^*(q_t)/\delta$ on average, which gives $\tau^*(q_t)\Delta x(q_t)$. Finally, a constrained environmentalist bears the adjustment cost $\bar{\tau}_e^2/(2\rho)$, whereas a materialist bears $\tau^*(q_t)^2/(2\rho)$. Since $\tau^*(q_t) \geq \bar{\tau}_e$ in the *BU* regime, the third term is non-negative.[19]

The last two terms are positive whatever the social valuation of environmental quality, which implies that our welfare results do not rely only on the fact that a greener population attaches more weight to a given environmental quality. Appendix D.5 considers a criterion in which this valuation is held fixed across cultural compositions. At the political-cultural steady state, where $\tau^*(q_H^*) = \tau^C$ and $\Delta x(q_H^*) = s$, removing the first term of (42) leaves $\tau^C s + [(\tau^C)^2 - \bar{\tau}_e^2]/(2\rho) > 0$, so that the results derived below also hold under this criterion.

5.2. The Cultural Cost of Carbon Pricing

One-period deviation. To identify this externality, we consider a one-period deviation from the political rule. At date t, the cultural composition is q_t and the carbon tax is set at an arbitrary level τ_t.

[19] In the *UU* regime, the same reasoning gives $d\mathcal{W}^P(q_t)/dq_t = (\sigma_e - \sigma_m)Q^P(q_t) + \tau^*(q_t)(\alpha_m - \alpha_e)$, since both types then bear the same adjustment cost $\tau^*(q_t)^2/(2\rho)$. Current welfare along the political rule is therefore also increasing in q_t in this regime whenever $Q^P(q_t) \geq 0$.

From date $t+1$ onward, the tax is set according to the political rule. Discounted welfare is then given by

$$\Psi(\tau_t; q_t) \equiv \mathcal{W}(\tau_t, q_t) + \beta J^P\big(H(q_t; \tau_t)\big). \tag{43}$$

The first term is current welfare; the second is the continuation welfare of the next generation, whose cultural composition $q_{t+1} = H(q_t; \tau_t)$ depends on the current tax. Differentiating (43) with respect to τ_t yields

$$\frac{\partial \Psi(\tau_t; q_t)}{\partial \tau_t} = \frac{\partial \mathcal{W}(\tau_t, q_t)}{\partial \tau_t} - \Omega(\tau_t; q_t), \tag{44}$$

where

$$\Omega(\tau_t; q_t) \equiv -\beta \frac{dJ^P(q_{t+1})}{dq_{t+1}} \times \frac{\partial H(q_t; \tau_t)}{\partial \tau_t}, \qquad \text{with } q_{t+1} = H(q_t; \tau_t). \tag{45}$$

The first term of (44) is the marginal effect of the tax on current welfare: the marginal environmental benefit of the tax net of the marginal adjustment cost it imposes on current households. The second term, $\Omega(\tau_t; q_t)$, is what we call the *marginal cultural cost of the carbon tax*. It measures the welfare loss that a marginal increase in the current tax imposes on future generations, through the cultural composition of the next generation. This cost is the product of two terms: the effect of the current tax on the share of environmentalists in the next generation, $\partial H/\partial \tau_t$, and the value of this share for future welfare, dJ^P/dq_{t+1}.

The political tax only accounts for the first term. Since $\partial \mathcal{W}(\tau_t, q_t)/\partial \tau_t = 0$ at $\tau_t = \tau^*(q_t)$, the welfare effect of a marginal increase in the political tax reduces to

$$\left.\frac{\partial \Psi(\tau_t; q_t)}{\partial \tau_t}\right|_{\tau_t = \tau^*(q_t)} = -\Omega(q_t), \qquad \text{where } \Omega(q_t) \equiv \Omega\big(\tau^*(q_t); q_t\big). \tag{46}$$

When $\Omega(q_t) > 0$, a marginal reduction of the tax below its political level increases discounted welfare: the political tax is too high, because voters do not take its cultural cost into account.[20] We now study the two terms of $\Omega(\tau_t; q_t)$ in turn, before showing how $\Omega(q_t)$ evolves with the cultural composition.

Effect of the tax on cultural transmission. Differentiating (38) with respect to τ_t gives

$$\frac{\partial H(q_t; \tau_t)}{\partial \tau_t} = 2q_t(1-q_t)\,\phi(\Delta x(\tau_t) - s)\,\frac{d\Delta x(\tau_t)}{d\tau_t}.$$

In the UU regime, the behavioral gap is equal to $\alpha_m - \alpha_e$ whatever the tax, so that $\partial H/\partial \tau_t = 0$ and the carbon tax has no cultural cost. In the BU regime, by contrast, a higher tax only affects materialists, so that $d\Delta x(\tau_t)/d\tau_t = -1/\rho$ and

$$\frac{\partial H(q_t; \tau_t)}{\partial \tau_t} = -\frac{2q_t(1-q_t)}{\rho}\,\phi(\Delta x(\tau_t) - s) < 0. \tag{47}$$

A higher current tax then reduces the share of environmentalists in the next generation. The mechanism is the one described in Section 3: by moving materialists towards environmentalists, a higher tax reduces the behavioral gap observed by children exposed to mixed influences, whose mass is $2q_t(1-$

[20] For $q_t > \bar{q}$, Ψ increases with τ_t below $\bar{\tau}_e$, since the marginal cultural cost is then equal to zero while $\partial \mathcal{W}(\tau_t, q_t)/\partial \tau_t > 0$. Under the conditions of Corollary 1, the tax that maximizes Ψ therefore lies between $\bar{\tau}_e$ and $\tau^*(q_t)$. It is equal to $\bar{\tau}_e$, the highest tax that leaves environmentalists unconstrained, when the marginal cultural cost exceeds $\partial \mathcal{W}/\partial \tau_t$ over the whole BU regime.

q_t), and therefore the probability that they adopt the environmental trait. The change in behavioral incidence studied in Section 4 thus generates a cultural externality: when they choose the tax, current voters do not take into account its effect on the cultural composition of the next generation.

Value of the cultural composition. Consider now the marginal value of the cultural composition at the political-cultural steady state q_H^*, where $G(q_H^*) = q_H^*$, $\tau^*(q_H^*) = \tau^C$ and $\Delta x(q_H^*) = s$. Differentiating the recursive relation $J^P(q_t) = \mathcal{W}^P(q_t) + \beta J^P(G(q_t))$ at $q_t = q_H^*$ yields

$$\frac{dJ^P(q_H^*)}{dq} = \frac{d\mathcal{W}^P(q_H^*)}{dq} + \beta\, G'(q_H^*)\, \frac{dJ^P(q_H^*)}{dq},$$

that is,

$$\frac{dJ^P(q_H^*)}{dq} = \frac{1}{1-\beta G'(q_H^*)}\, \frac{d\mathcal{W}^P(q_H^*)}{dq}, \qquad \text{with } G'(q_H^*) = 1 - \frac{2q_H^*(1-q_H^*)\,\phi(0)\,\delta(\sigma_e - \sigma_m)}{\rho}. \tag{48}$$

An additional environmentalist at q_H^* raises current welfare by $d\mathcal{W}^P/dq$, and this gain is propagated to future generations through the cultural dynamics, which explains the multiplier $1/(1-\beta G'(q_H^*))$. When q_H^* is locally stable, that is, when (32) holds, $|G'(q_H^*)| < 1$, so that this multiplier is positive and $dJ^P(q_H^*)/dq$ has the same sign as $d\mathcal{W}^P(q_H^*)/dq$.

The political tax is too high. From there, combining (46), (47) and (48), we can claim that:

Proposition 3 (Excessive political carbon tax)**.** *Suppose Assumptions 2 and 3 hold, q_H^* is locally stable, and $d\mathcal{W}^P(q_H^*)/dq > 0$. For every $\beta \in (0,1)$, there exists $\bar{\varepsilon} > 0$ such that, for every $\varepsilon \in (0, \bar{\varepsilon})$, setting $\tau_t = \tau^C - \varepsilon$ for one period and then returning to the political rule yields strictly higher discounted welfare than setting $\tau_t = \tau^C$.*

Proof. See Appendix D.2. □

Proposition 3 shows that, at the political-cultural steady state, the political tax is too high from an intertemporal point of view. To understand this result, consider the two effects of a reduction of the tax by ε. On the one hand, from (37), current welfare is a concave quadratic function of τ_t, maximized at τ^C when $q_t = q_H^*$. The current generation therefore loses

$$\mathcal{W}(\tau^C, q_H^*) - \mathcal{W}(\tau^C - \varepsilon, q_H^*) = (1 - q_H^*)\frac{\varepsilon^2}{2\rho}.$$

Since τ^C maximizes current welfare, this loss is proportional to the square of ε: it is negligible when the reduction is small. On the other hand, the reduction raises the behavioral gap from s to $s + \varepsilon/\rho$, and therefore the share of environmentalists in the next generation by approximately $2q_H^*(1-q_H^*)\phi(0)\varepsilon/\rho$. Valued at the marginal value of the cultural composition (48), this generates a future gain proportional to ε. For a small reduction, the welfare effect of the deviation is therefore approximately

$$\Psi(\tau^C - \varepsilon; q_H^*) - \Psi(\tau^C; q_H^*) \simeq \frac{2\beta q_H^*(1-q_H^*)\phi(0)}{\rho}\, \frac{dJ^P(q_H^*)}{dq}\, \varepsilon - (1-q_H^*)\frac{\varepsilon^2}{2\rho}, \tag{49}$$

which is positive as long as $\varepsilon < 4\beta q_H^* \phi(0)\, dJ^P(q_H^*)/dq$. The intuition is the following. Since the

political tax maximizes current welfare, a small reduction of this tax has almost no cost for the current generation, whereas it increases the behavioral gap, and therefore the share of environmentalists in the next generation, in proportion to the reduction. A gain proportional to ε always outweighs a loss proportional to ε^2 when ε is small enough.

This proposition exhibits a welfare-improving deviation; it does not characterize the optimal policy path. Its message is that a sequence of taxes that are Pigouvian for each generation is not necessarily efficient from an intertemporal point of view once current policy affects the future cultural composition. Moreover, since the cultural cost of the tax only exists in the *BU* regime, our result is not that carbon taxes are too high in general, but that the same Pigouvian rule becomes dynamically excessive once environmentalists are constrained and additional abatement is supplied by materialists.

Beyond the steady state. This result can be extended to every cultural composition above $\bar{q}$ when G is non-decreasing on $[\bar{q}, 1)$. Since $G(\bar{q}) > \bar{q}$, G then maps $[\bar{q}, 1)$ into itself. If, in addition, $Q^P(q_t) \geq 0$ on this interval, $\mathcal{W}^P$ is strictly increasing by Lemma 5, and so is J^P, as a discounted sum of increasing functions of q_t. Equations (45), (46) and (47) then yield the following corollary.

Corollary 1 (Cultural cost before the cultural reversal)**.** *Suppose that G is non-decreasing and $Q^P(q_t) \geq 0$ on $[\bar{q}, 1)$. Then, for every $q_t \in (\bar{q}, 1)$, $\Omega(q_t) > 0$: a marginal reduction of the current tax below $\tau^*(q_t)$ increases discounted welfare. In particular, this holds for $\bar{q} < q_t < q_H^*$, where environmental values are still spreading.*

Corollary 1 shows that the cultural cost of carbon pricing appears as soon as behavioral convergence begins, and not only once environmental values start to decline. Over the range $(\bar{q}, q_H^*)$, environmental values are still spreading while current policy already reduces future welfare by weakening their transmission. To illustrate this point, we make the marginal cultural cost explicit along the political-equilibrium path, where $\tau_t = \tau^*(q_t)$ and $q_{t+1} = G(q_t)$. From (47) and (46), we obtain

$$\Omega(q_t) = \begin{cases} 0, & q_t < \bar{q}, \\ \dfrac{2\beta q_t(1-q_t)}{\rho}\, \phi(\Delta x(q_t) - s)\, \dfrac{dJ^P(q_{t+1})}{dq_{t+1}}, & \bar{q} \leq q_t < 1, \end{cases} \qquad \text{with } q_{t+1} = G(q_t). \tag{50}$$

Figure 4 depicts $\Omega(q_t)$, that is, for each cultural composition, the gain in discounted welfare that a marginal reduction of the political tax would generate.

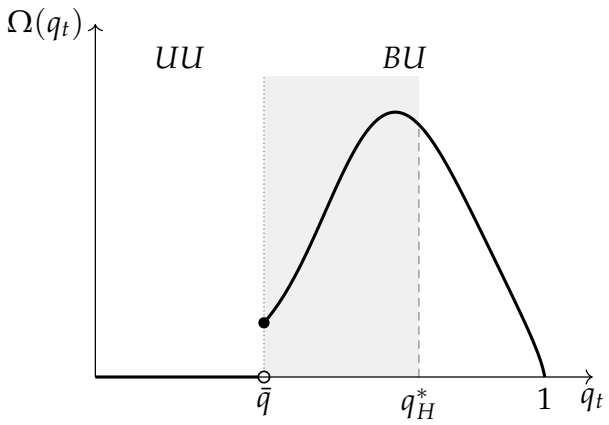


Figure 4: The marginal cultural cost of the carbon tax along the political-equilibrium path. The shaded area corresponds to the range of cultural compositions, $(\bar{q}, q_H^*)$, in which environmental values are still spreading.

Equation (50) and Figure 4 highlight three properties of the marginal cultural cost. First, $\Omega(q_t)$ is equal to zero in the UU regime and jumps to a strictly positive value at $\bar{q}$. Indeed, when q_t tends to $\bar{q}$ from above, $\Delta x(q_t)$ tends to $\alpha_m - \alpha_e = 2s$, so that $\Omega(q_t)$ tends to $2\beta\bar{q}(1-\bar{q})\phi(s)\left[dJ^P/dq\right]/\rho > 0$ under the conditions of Corollary 1. The cultural cost thus appears as soon as environmentalists stop adjusting to the tax, rather than increasing gradually with the level of the tax. Second, $\Omega(q_t)$ is strictly positive on $(\bar{q}, q_H^*)$, where the share of environmentalists is still increasing, so that a policymaker observing a growing environmentalist constituency would have no reason to think that current policy is dynamically excessive. Third, $\Omega(q_t)$ tends to zero when q_t tends to one. When almost every child has two environmentalist role models, the share of children exposed to mixed influences, $2q_t(1-q_t)$, vanishes, and the behavioral gap no longer affects cultural transmission. The exact shape of $\Omega(q_t)$ in the BU regime, and in particular the location of its maximum, depends on parameters.

These results also clarify the difference between our framework and Besley and Persson (2023). In their framework, greener economic conditions increase the relative fitness of green values, and limited commitment can make a welfare-improving green transition too slow. In our framework, once behavioral convergence begins, the welfare-improving deviation goes in the opposite direction: a lower current tax preserves the behavioral difference through which environmental values are transmitted. This difference does not come from lower environmental damages, but from the fact that the economic fitness of a cultural trait and its behavioral transmissibility can move in opposite directions: as shown in Lemma 3, the relative payoff of environmentalists keeps increasing in the BU regime, while their transmission advantage declines.

5.3. Commitment and Long-Run Welfare

Proposition 3 is a local result. We now compare complete policy paths. Consider a committed policy in which the carbon tax is set at a constant level $\widehat{\tau} \in (\bar{\tau}_e, \tau^C)$ at every date. We denote by $\widehat{q}_t$, $\widehat{X}_t$, $\widehat{Q}_t$ and $\widehat{\mathcal{V}}(q_0)$ the cultural composition, aggregate carbon use, environmental quality and discounted welfare along this path. Variables along the political path are denoted with a superscript P.

Long-run outcomes under commitment. Under commitment, environmentalists are constrained, since $\widehat{\tau} > \bar{\tau}_e$, but the behavioral gap, $\alpha_m - \underline{x} - \widehat{\tau}/\rho$, remains above s, since $\widehat{\tau} < \tau^C$. From (28), environmental values therefore keep a strictly positive fitness advantage whatever the cultural composition, and the whole population eventually becomes environmentalist: $\widehat{q}_t \to 1$. In the long run, every household consumes $\underline{x}$ and bears the adjustment cost $\bar{\tau}_e^2/(2\rho)$. From (37), we obtain

$$\widehat{X}^{\infty} = \underline{x}, \qquad \widehat{Q}^{\infty} = \bar{Q} - \delta\underline{x}, \qquad \widehat{\mathcal{W}}^{\infty} = w + \sigma_e\left(\bar{Q} - \delta\underline{x}\right) - \frac{\bar{\tau}_e^2}{2\rho}. \tag{51}$$

Long-run outcomes under the political rule. Under the political rule, any initial composition in the basin of attraction of q_H^* leads to $q_t^P \to q_H^*$ and $\tau_t^* \to \tau^C$. At this steady state, environmentalists consume $\underline{x}$, while materialists consume $\underline{x} + s$ and bear the adjustment cost $(\tau^C)^2/(2\rho)$. From (40), we obtain

$$X^{P,\infty} = \underline{x} + (1 - q_H^*)s, \qquad Q^{P,\infty} = \bar{Q} - \delta\underline{x} - \delta(1 - q_H^*)s, \tag{52}$$

and

$$\mathcal{W}^{P,\infty} = w + \left[q_H^*\sigma_e + (1 - q_H^*)\sigma_m\right] Q^{P,\infty} - q_H^*\frac{\bar{\tau}_e^2}{2\rho} - (1 - q_H^*)\frac{(\tau^C)^2}{2\rho}. \tag{53}$$

Long-run welfare gap. Subtracting (53) from (51), and using $\delta\left[q_H^*\sigma_e + (1 - q_H^*)\sigma_m\right] = \tau^C$, we obtain

$$\widehat{\mathcal{W}}^{\infty} - \mathcal{W}^{P,\infty} = (1 - q_H^*)\left[(\sigma_e - \sigma_m)(\bar{Q} - \delta\underline{x}) + \tau^C s + \frac{(\tau^C)^2 - \bar{\tau}_e^2}{2\rho}\right]. \tag{54}$$

This gap is the welfare effect of replacing the $(1 - q_H^*)$ materialists of the political steady state by environmentalists, and its three terms have the same interpretation as in (42): a higher valuation of environmental quality, a lower carbon use, and lower adjustment costs. Moreover, since $\bar{Q} - \delta\underline{x} = Q^{P,\infty} + \delta(1 - q_H^*)s > Q^{P,\infty}$, the term in brackets is strictly larger than $d\mathcal{W}^P(q_H^*)/dq$, so that long-run welfare is strictly higher under commitment whenever $d\mathcal{W}^P(q_H^*)/dq > 0$. Since discounted welfare is mainly determined by its long-run level when β is close to one, we can then claim that:

Proposition 4 (Commitment and long-run welfare)**.** *Suppose Assumptions 2 and 3 hold, q_H^* is locally stable, and $d\mathcal{W}^P(q_H^*)/dq > 0$. For any $\widehat{\tau} \in (\bar{\tau}_e, \tau^C)$ and any initial condition q_0 in the basin of attraction of q_H^*, there exists $\bar{\beta}(q_0, \widehat{\tau}) < 1$ such that, for all $\beta > \bar{\beta}(q_0, \widehat{\tau})$,*

$$\widehat{\mathcal{V}}(q_0) > \mathcal{V}^P(q_0).$$

Moreover, long-run carbon use is strictly lower under commitment: $\widehat{X}^{\infty} < X^{P,\infty}$.

Proposition 4 shows that a lower committed carbon tax can generate both higher welfare and lower carbon use in the long run than the sequence of taxes chosen by successive electorates. The committed tax may generate less abatement in the short run, but this ranking is reversed in the long run, because the current tax shapes the cultural composition of future generations and, through it, future carbon demand. This result does not mean that carbon taxes should be lower in general. In the *UU* regime, there is no dynamic wedge, and a tax high enough to constrain both cultural types removes the behavioral margin altogether. The result only concerns the intermediate range in which environmentalists are constrained while materialists still respond to the tax.

5.4. Environmental Ranking and Preference Robustness

The criterion (39) evaluates the welfare of each generation according to its own preferences. Since the cultural composition is endogenous, part of the welfare difference between two paths may come from changes in the weight that the population attaches to environmental quality. To address this issue, we consider a criterion that does not compare utilities across cultural compositions. For a policy path π, we define discounted environmental quality as

$$\mathcal{E}^{\pi}(q_0) = \sum_{t=0}^{\infty} \beta^t Q_t^{\pi}. \tag{55}$$

Under this criterion, the cultural composition has no social value per se. From (51) and (52), the long-run difference in environmental quality between the two paths is

$$\widehat{Q}^{\infty} - Q^{P,\infty} = \delta(1 - q_H^*)s > 0. \tag{56}$$

This difference has a simple interpretation: in the long run, the political path keeps a share $1 - q_H^*$ of materialists, each of whom consumes s more of the carbon-intensive good than an environmentalist. Then, we can claim that:

Corollary 2 (Preference-invariant environmental ranking)**.** *Suppose Assumptions 2 and 3 hold, q_H^* is locally stable, and q_0 belongs to its basin of attraction under the political rule. For any $\widehat{\tau} \in (\bar{\tau}_e, \tau^C)$, there exists $\bar{\beta}_E(q_0, \widehat{\tau}) < 1$ such that, for all $\beta > \bar{\beta}_E(q_0, \widehat{\tau})$,*

$$\widehat{\mathcal{E}}(q_0) > \mathcal{E}^P(q_0).$$

Contrary to Proposition 4, Corollary 2 does not require any condition on $d\mathcal{W}^P(q_H^*)/dq$. It only relies on the cultural dynamics and on the resulting long-run gap in carbon use. In addition, Appendix D.5 considers an intermediate criterion, in which environmental quality has the same social value whatever the cultural composition, while actual adjustment costs are taken into account. Both the cultural cost of carbon pricing and the long-run welfare ranking hold under this criterion. Our results therefore do not rely on considering a larger environmentalist population as desirable in itself.

To sum up, a carbon tax that is efficient from a static point of view is not necessarily efficient from an intertemporal point of view when policy affects the behavior through which preferences are transmitted. Carbon pricing can keep reducing current emissions while weakening the transmission of the environmental values that lower future carbon demand. The fact that the economic fitness of environmental values and their behavioral transmissibility can move in opposite directions is the source of the welfare results of this section.

6. Purposeful Parental Socialization

In the baseline model, environmental values are transmitted through social learning. A child exposed to mixed influences adopts the environmental trait only if the behavior of her role models provides sufficient evidence that this trait shapes behavior, and she observes this behavior without observing the constraints that shape it. One may then wonder whether the weakening of cultural transmission established in Proposition 1 rests on this attribution problem. Here, we show that it does not. We

replace social learning by purposeful parental socialization, in the spirit of Bisin and Verdier (2001), and leave the economic and political blocks of the model unchanged: for any inherited share q_t, households make the same consumption choices and, since adults still vote according to their own current utility, the carbon tax is still given by Lemma 2. Parents now perfectly understand how the carbon tax and the lower bound shape the behavior of their child. Nevertheless, behavioral convergence weakens the transmission of environmental values, because it reduces the difference that these values make to the behavior of the child, and therefore the return to transmitting them.

This extension builds on two contributions. First, as in models of green socialization (Bezin, 2015, 2019), materialism is the default trait, and environmental values have to be transmitted through costly socialization. In Bezin (2019), the incentive to transmit green preferences responds to the economic environment, since relative prices and clean technologies affect the value that a green parent attaches to having a green child. Second, as in Carvalho (2013), the return to socialization depends on the behavioral difference between the two traits: in his model, the regulation of veiling changes the difference between the behavior of religious and secular types, and with it the incentive to invest in religious education. Our contribution is to make this behavioral difference an outcome of the political equilibrium. The cultural composition of one generation determines the carbon tax, which in turn changes the behavioral consequences of the trait that parents seek to transmit to the next generation.

We first characterize the socialization choices of parents (Section 6.1). We then study the resulting cultural dynamics (Section 6.2) and discuss what this extension reveals about our mechanism (Section 6.3).

6.1. Socialization Choices

Socialization technology. At the end of period t, each adult has one child and chooses an effort $a_t^i \geq 0$ to transmit environmental values to this child, where $i \in \{e, m\}$ denotes the type of the parent. This effort is costly and raises the probability that the child acquires the environmental trait. A child who does not acquire this trait becomes a materialist. Formally, the probability that the child becomes an environmentalist and the cost of socialization are respectively given by

$$T(a_t^i) = \mu + \gamma a_t^i \qquad \text{and} \qquad C(a_t^i) = \frac{\kappa}{2}\left(a_t^i\right)^2, \tag{57}$$

with $0 < \mu < 1$, $\gamma > 0$, $\kappa > 0$ and $a_t^i \leq (1-\mu)/\gamma$, such that $T(a_t^i)$ is a probability.[21] The parameter μ is the probability that a child acquires environmental values in the absence of any parental effort, for instance through school, peers or the media, γ measures the effectiveness of parental socialization, and κ its cost. The two types of parents have access to the same technology, and this technology can only raise the probability of the environmental trait. As in the baseline model, materialism is therefore the default orientation of the child.

Parental preferences. Following Bisin and Verdier (2001), parents are paternalistic and exhibit imperfect empathy: they anticipate the behavior that their child will adopt given her cultural trait, but evaluate this behavior with their own preferences. Moreover, parents are myopic: they evaluate the situation of their child under the current economic and political environment, that is, for given q_t,

[21] Assumption 4 below ensures that this upper bound never binds.

carbon tax τ_t^*, environmental quality Q_t and transfer R_t.[22] Following the same steps as for the indirect utility (15), the value that a parent of type i attaches to a child of type $j \in \{e, m\}$ is

$$V_{ij}(q_t) = w + R_t - \tau_t^* x_t^j - \frac{\rho}{2}\left(x_t^j - \alpha_i\right)^2 + \sigma_i Q_t, \tag{58}$$

where $x_t^j = \max\{\underline{x}, \alpha_j - \tau_t^*/\rho\}$ is the carbon-intensive consumption of a type-j adult, given by (5) at $\tau_t = \tau_t^*$. The first index refers to the preferences used to evaluate the child and the second to the behavior generated by her cultural trait, such that $V_{ii}(q_t)$ coincides with the equilibrium payoff $V_i(q_t)$ defined in Section 2.3. Since an individual child is atomistic, her trait affects neither R_t, nor Q_t, nor the political equilibrium. The gain that a type-i parent derives from having an environmentalist rather than a materialist child is entirely determined by the behavior of the child. Using the fact that, for any x,

$$-\tau_t^* x - \frac{\rho}{2}(x - \alpha_i)^2 = \frac{(\tau_t^*)^2}{2\rho} - \alpha_i \tau_t^* - \frac{\rho}{2}\left(x - \alpha_i + \frac{\tau_t^*}{\rho}\right)^2,$$

this gain can be written as

$$V_{ie}(q_t) - V_{im}(q_t) = \frac{\rho}{2}\left[\left(x_t^m - \alpha_i + \frac{\tau_t^*}{\rho}\right)^2 - \left(x_t^e - \alpha_i + \frac{\tau_t^*}{\rho}\right)^2\right]. \tag{59}$$

The level $\alpha_i - \tau_t^*/\rho$ is the carbon-intensive consumption that a type-i parent would like her child to choose under the current tax, ignoring the lower bound. In other words, socialization is valuable to a parent insofar as it brings the behavior of her child closer to the behavior she would choose for her.

Let us highlight that this gain differs from the relative payoff of environmentalists, $D(q_t) = V_{ee}(q_t) - V_{mm}(q_t)$, studied in Section 2.3. The latter compares two adults, each of them evaluated with her own preferences. Here, the preferences of the parent are held fixed, and only the behavior of the child varies. As we will see, the economic attractiveness of environmental values and the incentive to transmit them can move in opposite directions.

Environmentalist parents. Consider first an environmentalist parent. A materialist child consumes $x_t^m = \alpha_m - \tau_t^*/\rho$, that is, $\bar{r}$ more than the level preferred by her parent, whatever the tax. An environmentalist child consumes exactly the level preferred by her parent as long as she is unconstrained. In the *BU* regime, however, she is constrained at $\underline{x}$, which exceeds the level preferred by her parent by $\underline{x} - \alpha_e + \tau_t^*/\rho = (\tau_t^* - \bar{\tau}_e)/\rho$. This distance is precisely the part of the initial behavioral gap that has been closed by the tax, $\bar{r} - \Delta x(q_t)$: beyond $\bar{\tau}_e$, any further increase in the tax reduces the consumption of a materialist child, but no longer that of an environmentalist child. Substituting into (59), we obtain, in both regimes,

$$V_{ee}(q_t) - V_{em}(q_t) = \frac{\rho}{2}\left\{\bar{r}^2 - [\bar{r} - \Delta x(q_t)]^2\right\}. \tag{60}$$

[22] The same assumption is made in Bezin (2019). It rules out the multiple equilibria that may arise when parents form expectations about the future cultural composition, which itself depends on their socialization choices (see Carvalho, 2013). It is also consistent with the political block of the model, in which voters do not take into account the effect of the carbon tax on the next generation.

From (23) and $\bar{\tau}_e = \tau^*(\bar{q})$, the behavioral convergence induced by the tax in the BU regime is $\bar{r} - \Delta x(q_t) = \delta(\sigma_e - \sigma_m)(q_t - \bar{q})/\rho$, such that

$$V_{ee}(q_t) - V_{em}(q_t) = \begin{cases} \dfrac{\rho}{2}\bar{r}^2, & q_t < \bar{q}, \\ \dfrac{\rho}{2}\bar{r}^2 - \dfrac{\delta^2(\sigma_e - \sigma_m)^2}{2\rho}\left(q_t - \bar{q}\right)^2, & \bar{q} \leq q_t < 1. \end{cases} \tag{61}$$

In the UU regime, the carbon tax shifts by the same amount the level preferred by the parent and the behavior of both possible children. The behavioral difference generated by the cultural trait of the child remains equal to $\bar{r}$, and so does the return to environmental socialization. In the BU regime, the tax keeps moving the behavior of a materialist child, but no longer that of an environmentalist child. Successful socialization then reduces the carbon-intensive consumption of the child by $\Delta x(q_t) < \bar{r}$ only, and its value to an environmentalist parent declines. Put differently, carbon pricing now does part of the job that environmental socialization would otherwise do. Note that this decline is slow at first and accelerates as q_t increases: at $\bar{q}$, the lower bound coincides with the level preferred by the parent, such that a small departure from this level has a negligible cost.

Materialist parents. Consider now a materialist parent. The level she would like her child to choose, $\alpha_m - \tau_t^*/\rho$, is precisely the behavior of a materialist child, whereas an environmentalist child consumes $\Delta x(q_t)$ less. Equation (59) then gives

$$V_{me}(q_t) - V_{mm}(q_t) = -\frac{\rho}{2}\left[\Delta x(q_t)\right]^2 < 0. \tag{62}$$

A materialist parent always prefers a materialist child. Behavioral convergence weakens this preference, since an environmentalist child behaves more and more like a materialist one, but never reverses it.

Optimal socialization. A parent of type i chooses a_t^i to solve

$$\max_{a_t^i \geq 0} \left\{ T(a_t^i) V_{ie}(q_t) + \left[1 - T(a_t^i)\right] V_{im}(q_t) - \frac{\kappa}{2}\left(a_t^i\right)^2 \right\}. \tag{63}$$

The objective is strictly concave in a_t^i. When it is interior, the optimal effort satisfies the first-order condition

$$\gamma\left[V_{ie}(q_t) - V_{im}(q_t)\right] = \kappa a_t^i.$$

The left-hand side is the marginal benefit of socialization, that is, the effect of effort on the probability of having an environmentalist child times the gain associated with such a child. The right-hand side is its marginal cost. Using (60) and (62), we obtain

$$a_t^{e*} = \frac{\gamma}{\kappa}\left[V_{ee}(q_t) - V_{em}(q_t)\right] = \frac{\gamma\rho}{2\kappa}\left\{\bar{r}^2 - \left[\bar{r} - \Delta x(q_t)\right]^2\right\} \qquad \text{and} \qquad a_t^{m*} = 0. \tag{64}$$

Materialist parents would like to reduce the probability that their child becomes an environmentalist, which the socialization technology does not allow: the non-negativity constraint binds, and they exert no effort. As in Carvalho (2013), where secular parents choose zero religious education, both types face the same technology and the same cost, and the difference between their choices only comes from

the preferences with which they evaluate the behavior of their child.[23] From there, combining (61) and (64), we can claim that:

Lemma 6. *Suppose Assumption 2 holds. Materialist parents exert no socialization effort, $a_t^{m*} = 0$. Environmentalist parents exert a strictly positive effort, which is constant for $q_t < \bar{q}$ and strictly decreasing in q_t for $\bar{q} \leq q_t < 1$:*

$$a_t^{e*} = \begin{cases} \dfrac{\gamma\rho}{2\kappa}\bar{r}^2, & q_t < \bar{q}, \\ \dfrac{\gamma\rho}{2\kappa}\bar{r}^2 - \dfrac{\gamma\delta^2(\sigma_e - \sigma_m)^2}{2\kappa\rho}\left(q_t - \bar{q}\right)^2, & \bar{q} \leq q_t < 1. \end{cases} \tag{65}$$

Effort remains strictly positive in the *BU* regime because, under Assumption 2, materialists never reach the lower bound, such that $\Delta x(q_t) > 0$ for all $q_t < 1$. Lemma 6, illustrated in Figure 5(a), shows that the weakening of cultural transmission does not require children to misinterpret the behavior they observe. Environmentalist parents perfectly understand that the convergence of materialists results from the carbon tax. Yet, once environmentalists are constrained, the tax makes the cultural trait of the child less consequential for her behavior, and the private return to transmitting environmental values falls.

6.2. Cultural Dynamics

Denoting again by p_t^{ij} the probability that a child whose parent has trait $i \in \{e, m\}$ adopts trait $j \in \{e, m\}$ on date t, the socialization choices characterized in Lemma 6 yield the following transition probabilities:

$$\begin{aligned} p_t^{ee} &= \mu + \gamma a_t^{e*}, \quad & p_t^{em} &= 1 - \mu - \gamma a_t^{e*}, \\ p_t^{me} &= \mu, & p_t^{mm} &= 1 - \mu. \end{aligned} \tag{66}$$

The child of an environmentalist parent acquires environmental values with probability $\mu + \gamma a_t^{e*}$, whereas the child of a materialist parent only acquires them with the baseline probability μ. The share of environmentalists in the next generation is then given by

$$q_{t+1} = q_t p_t^{ee} + (1 - q_t) p_t^{me} = \mu + \gamma a_t^{e*} q_t \equiv \Gamma(q_t). \tag{67}$$

Let $\theta \equiv \gamma^2\rho\bar{r}^2/(2\kappa)$ denote the increase in the probability of transmission generated by an environmentalist parent when both cultural types are unconstrained. Using Lemma 6, the mapping Γ is given by

$$\Gamma(q_t) = \begin{cases} \mu + \theta q_t, & q_t < \bar{q}, \\ \mu + \theta q_t - \dfrac{\gamma^2\delta^2(\sigma_e - \sigma_m)^2}{2\kappa\rho} q_t \left(q_t - \bar{q}\right)^2, & \bar{q} \leq q_t < 1. \end{cases} \tag{68}$$

Below $\bar{q}$, a larger environmentalist share raises the number of parents who invest in environmental socialization, but leaves the effort of each of them unchanged, such that Γ is linear in q_t. Above $\bar{q}$, a

[23] If materialist parents could also invest in the transmission of their own trait, as in Bisin and Verdier (2001), behavioral convergence would reduce their incentive to do so as well, since their gain from having a materialist child, $\rho[\Delta x(q_t)]^2/2$, falls as the behavioral gap shrinks. The weakening of environmental socialization established below would remain, but the effect of behavioral convergence on the cultural mapping would have the sign of $q_t\bar{r} - \Delta x(q_t)$. Close to $\bar{q}$, the incentive of materialist parents would decline faster than that of environmentalist parents, and convergence would initially favor environmental values. The decline in the steady-state share of environmentalists established in Proposition 5 therefore relies on the directed structure of socialization.

larger environmentalist share still raises the number of environmentalist parents, but it also raises the carbon tax, which reduces the behavioral gap and, with it, the effort of each environmentalist parent. The last term of (68) measures the resulting loss in cultural transmission. It is equal to zero at $\bar{q}$ and increases with q_t. Its magnitude is governed by $\delta(\sigma_e - \sigma_m)/\rho$, the rate at which a larger environmentalist share raises the carbon tax and reduces the behavioral gap (Lemma 3): the more responsive policy is to the cultural composition of the electorate, the faster behavioral convergence proceeds, and the larger this loss.

To isolate the role of behavioral convergence, we compare Γ with the mapping that would prevail if the behavioral gap were fixed to $\bar{r}$, as in Figure 3:

$$\Gamma^{\bar{r}}(q_t) \equiv \mu + \theta q_t. \tag{69}$$

The two mappings coincide up to $\bar{q}$, and Γ lies strictly below $\Gamma^{\bar{r}}$ beyond. Rearranging (67), the law of motion of q_t can be written as

$$q_{t+1} - q_t = (1 - q_t)\mu - q_t\left(1 - \mu - \gamma a_t^{e*}\right). \tag{70}$$

The first term is the inflow into environmentalism, made of the children of materialist parents who acquire environmental values. The second term is the outflow from environmentalism, made of the children of environmentalist parents who do not acquire them. As q_t rises, the inflow falls, since there are fewer materialist parents. Moreover, beyond $\bar{q}$, environmentalist parents reduce their effort, such that the outflow rate, $1 - \mu - \gamma a_t^{e*}$, increases. We impose the following conditions.

Assumption 4.

$$\mu + \theta < 1 \qquad \textit{and} \qquad \mu > (1 - \theta)\bar{q}, \qquad \textit{with } \theta = \frac{\gamma^2 \rho \bar{r}^2}{2\kappa}.$$

The first condition ensures that the transmission probability $p_t^{ee} = \mu + \gamma a_t^{e*}$, which is at most equal to $\mu + \theta$, remains below one. The second condition can be written as $\Gamma(\bar{q}) > \bar{q}$: environmental values are still spreading when environmentalists reach the lower bound. Under these two conditions, the fixed-gap mapping (69) admits a unique steady state,

$$q_{PS}^{\bar{r}} = \frac{\mu}{1 - \theta} \in (\bar{q}, 1). \tag{71}$$

In other words, in the absence of behavioral convergence, environmental values would keep spreading beyond the share at which environmentalists exhaust their adjustment margin. Any reduction in the steady-state share of environmentalists can only be generated in the BU regime, rather than inherited from the dynamics of the UU regime. The properties of cultural dynamics are summarized in the following proposition.

Proposition 5. *Suppose Assumptions 2 and 4 hold.*

(i) There exists a unique steady state $q_{PS}^* \in (\bar{q}, 1)$. The environmentalist share increases for $0 \leq q_t < q_{PS}^*$ and decreases for $q_{PS}^* < q_t < 1$.

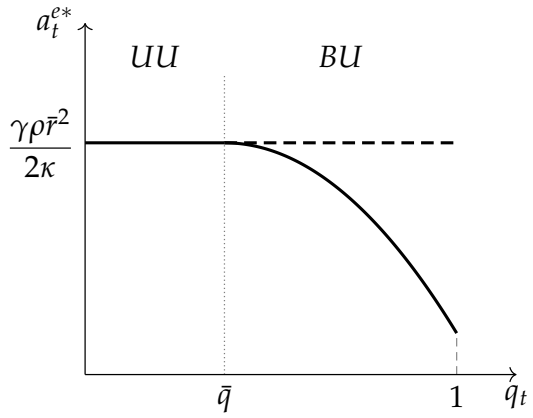
(ii) Behavioral convergence lowers the steady-state share of environmentalists: $q_{PS}^* < q_{PS}^{\bar{r}}$.

(iii) The steady state q^*_{PS} is locally stable if and only if

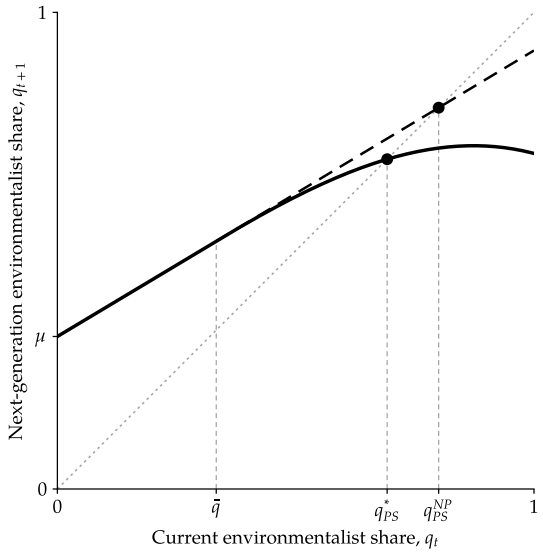
$$\frac{\gamma^2\delta^2(\sigma_e-\sigma_m)^2}{2\kappa\rho}\left(q^*_{PS}-\bar{q}\right)\left(3q^*_{PS}-\bar{q}\right)<1+\theta. \tag{72}$$

Proof. See Appendix E. □

The proof relies on the change in the environmentalist share, $\Gamma(q_t)-q_t$. This change is positive at $\bar{q}$ by Assumption 4, strictly decreasing in q_t in the BU regime, and negative as q_t tends to one, which gives a unique steady state. At $q^{\bar{r}}_{PS}$, the terms of (68) that determine the fixed-gap steady state cancel out, such that only the loss in cultural transmission remains. The change in the environmentalist share is negative at this point, which places q^*_{PS} below $q^{\bar{r}}_{PS}$. Finally, as in Proposition 1, the slope of the mapping at the steady state, $\Gamma'(q^*_{PS})$, is always below one, and exceeds -1 if and only if (72) holds. When the tax reacts strongly to the cultural composition, $\Gamma'(q^*_{PS})$ may become negative, and the environmentalist share then oscillates around q^*_{PS}.



(a) Socialization effort of environmentalist parents, a_t^{e*}.



(b) Dynamics of environmental values, $\Gamma(q_t)$.

Figure 5: Purposeful parental socialization. The solid curves depict the equilibrium effort and mapping. The dashed curves depict the effort and mapping obtained when the behavioral gap is fixed to $\bar{r}$.

Figure 5 illustrates the two steps behind Proposition 5. As long as $q_t<\bar{q}$, the cultural trait of the child generates the same behavioral difference, and environmentalist parents keep the same socialization effort (panel a). Beyond $\bar{q}$, the equilibrium carbon tax brings the behavior of a materialist child closer to that of an environmentalist child, and environmentalist parents invest less in socialization. The mapping Γ then falls below its fixed-gap counterpart, and crosses the 45° line at $q^*_{PS}<q^{\bar{r}}_{PS}$ (panel b). From (70), the steady state is reached when the inflow into environmentalism equals the outflow from it. Behavioral convergence raises the outflow rate at every $q_t>\bar{q}$, such that this balance is reached at a lower environmentalist share.

Proposition 5 parallels Proposition 1. Under both transmission processes, environmental values spread when both cultural types are unconstrained, the carbon tax weakens their transmission once

environmentalists reach the lower bound, and the steady state lies in the BU regime. The two results differ in the role played by behavioral convergence. Under social learning, environmental values would spread to the whole population if the behavioral gap remained equal to $\bar{r}$, and convergence is what generates an interior steady state. Under parental socialization, imperfect transmission within environmentalist families already generates an interior steady state, and convergence lowers it.

6.3. Discussion

Attractiveness versus transmission. This extension makes particularly clear that the weakening of cultural transmission is not driven by a deterioration in the relative position of environmentalists. As shown in Lemma 3, their relative payoff, $D(q_t)$, keeps increasing with their population share. At the same time, by (61), the value that an environmentalist parent attaches to transmitting her values decreases once environmentalists are constrained. Environmental values thus become more rewarding to hold, but less valuable to transmit. The reason is that $D(q_t)$ evaluates each trait with its own preferences, and therefore captures the gain from valuing environmental quality more. The return to socialization, by contrast, evaluates the behavior generated by each trait with the same preferences, those of the parent, and therefore only captures the behavioral difference between the two traits.

What the two transmission processes have in common. Under social learning, behavioral convergence makes the example of environmentalists less distinctive, such that children infer that environmental values have little behavioral content. Under parental socialization, it reduces the private return to investing in environmental education, although parents perfectly understand why materialists behave like environmentalists. What the two microfoundations share is therefore not the attribution problem, but the behavioral relevance of cultural values: in both cases, the transmission of values increases with the difference they make to behavior, and carbon pricing reduces this difference once environmentalists are constrained. The decline in the steady-state share of environmentalists established in Proposition 5 additionally relies on the directed structure of socialization, in which materialism is the default orientation of the child.

Relation to the literature. In Bezin (2019), economic conditions and clean technologies affect green parenting by changing the value associated with a green child, which generates complementarities between green culture and the economic environment. In a similar vein, Besley and Persson (2019b) and Besley and Persson (2023) show how environmental values, environmental policy and green technologies can reinforce one another over time. Our mechanism introduces a countervailing force. Once environmentalists have exhausted their adjustment margin, stronger carbon pricing induces materialists to adopt an increasing part of the low-carbon behavior otherwise associated with environmental values, and reduces the return to transmitting these values. Environmental policy and environmental culture need not remain complements throughout the transition. Relative to Carvalho (2013), the policy that changes the behavioral consequences of the transmitted trait is chosen by the current cultural composition of the electorate. A cultural group thereby affects, through the policy it supports, the transmission of its own values to the next generation.

Our mechanism also differs from cultural backlash. In Carvalho et al. (2024), the state transmits a mainstream trait through the education system, and strengthening this transmission can induce alternative groups to intensify their own socialization, to the point of reducing the long-run share of the

mainstream trait. No such defensive response is at work here. Materialist parents do not resist carbon pricing, and environmentalist parents do not reduce their effort because policy threatens their values. Carbon pricing instead reduces the behavioral consequences of successful environmental socialization, and thereby its return. The two mechanisms can thus generate similar countervailing cultural effects through opposite behavioral responses. In cultural-backlash models, stronger intervention can weaken transmission because the targeted groups resist it. Here, carbon pricing weakens transmission because the targeted groups comply with it.

This provides a precise interpretation of "compliance without conviction". Carbon pricing can reduce current carbon-intensive consumption by inducing materialists to behave increasingly like environmentalists, while weakening the transmission of the environmental values that support stringent climate policy. Whether these values are transmitted through social learning or through deliberate parental investment, the success of current policy and the transmission of the values that support it can move in opposite directions.

7. Conclusion

Standard analyses of environmental policy evaluate carbon pricing through its effect on current behavior. Recent work has added a dynamic dimension, by allowing environmental values themselves to evolve with economic and political conditions (Besley and Persson, 2019b; Bezin, 2019; Besley and Persson, 2023). In this literature, policy affects the diffusion of environmental values by changing their relative attractiveness. In this paper, we argue that this is not the only channel, and we propose a simple theoretical framework in which carbon pricing also changes the extent to which differences in values are expressed in differences in behavior. Since environmental values are partly transmitted through the behavior that distinguishes them, current climate policy then shapes the cultural composition, and therefore the political support for climate policy, inherited by future generations.

We conclude that the mutual interplay between environmental values and carbon pricing may become self-limiting. A greener electorate always chooses a more stringent carbon tax. As long as both cultural groups retain an adjustment margin, this tax reduces carbon use without affecting the behavioral difference between them. Once environmentalists have exhausted their adjustment margin, however, the same political rule obtains additional abatement from materialists only, whose behavior progressively converges towards that of environmentalists. The transmission of environmental values then weakens, and eventually reverses, although the relative payoff of environmentalists keeps increasing. In other words, carbon pricing does not make environmental values less attractive; it makes them less consequential for behavior. Materialist compliance thus makes environmentalist conviction harder to transmit.

Our contribution can be summarized in three main points. Firstly, the cultural effect of climate policy depends on its behavioral incidence, that is, on whose behavior adjusts, and not only on the size of current abatement. Two interventions generating the same reduction in carbon use can have opposite effects on the transmission of environmental values: a higher carbon tax obtains this reduction from materialists and compresses behavioral differences, whereas expanding low-carbon alternatives obtains it from environmentalists and widens these differences. Compensating households for the financial burden of carbon pricing, by contrast, leaves behavioral differences unchanged. Behavioral incidence is therefore distinct from the distributional incidence usually emphasized in the climate-policy literature: it does not concern how the costs of policy are shared, but how the

adjustment it induces is allocated across cultural groups. Secondly, a sequence of carbon taxes that are Pigouvian for each generation is not necessarily efficient from an intertemporal point of view. Once environmentalists are constrained, the carbon tax has a marginal cultural cost that current voters do not take into account, such that a marginal reduction of the political tax increases discounted welfare. A constant carbon tax set between $\bar{\tau}_e$ and τ^C, that is, below the long-run political tax, also generates lower carbon use in the long run and, for sufficiently patient social evaluation, higher discounted welfare. Importantly, this ranking does not rely on considering a larger environmentalist population as desirable in itself. Thirdly, the mechanism does not depend on a particular transmission process. When parents deliberately socialize their children, they perfectly understand why materialists behave like environmentalists. Yet, the return to transmitting environmental values falls with the behavioral difference that these values make.

These results qualify the dynamic complementarity between environmental culture and climate policy emphasized in the literature (Bezin, 2019; Besley and Persson, 2019b, 2023). In our framework, this complementarity prevails in the early phase of the transition, but beyond a threshold, carbon pricing substitutes for the behavior that environmental values would otherwise generate. This switch requires no change in preferences, in the political rule, or in the relative position of either cultural group. It also differs from cultural backlash (Carvalho et al., 2024): policy weakens the transmission of environmental values not because the targeted group resists it, but because it complies with it. Even in our optimistic model of politics, in which parties maximize the average utility of the current generation and the carbon tax is statically efficient, the political equilibrium may thus put society on a path with higher long-run carbon use.

Some predictions of our model deserve to be empirically investigated. In particular, price incentives should compress behavioral differences across motivation types, especially when the most motivated individuals have already exhausted their adjustment possibilities. Evidence that rewards disproportionately mobilize less motivated blood donors (Goette and Stutzer, 2020), and that financial incentives make recycling less informative about the motives behind it (Ling and Xu, 2021), is consistent with this margin. Our analysis adds a dynamic prediction: where environmental attitudes are transmitted through the behavior of parents and peers, a policy-induced reduction in behavioral distinctiveness should be followed by a weaker transmission of these attitudes. Panel data linking the attitudes and the carbon-intensive consumption of parents and children would make it possible to test this prediction. Moreover, we advocate evaluations of climate policies that look not only at the size of the behavioral response they induce, but also at the groups that supply it.

Finally, the framework we propose is sufficiently tractable to be extended in several dimensions. For instance, investments in low-carbon alternatives, such as public transportation or cleaner energy systems, could be chosen through the political process jointly with the carbon tax. Since such investments are persistent while the carbon tax is reconsidered by each generation, this extension would raise new questions about commitment and the sequencing of climate-policy instruments. Another dimension deserving more attention is the political behavior of parents. In the current version of the model, adults vote according to their current utility. If they also cared about the cultural orientation of their children, environmentalists would support a lower carbon tax, so as to preserve the behavioral distinctiveness of their values, whereas materialists would support a higher one. The political equilibrium would then partly internalize the cultural cost of carbon pricing, through opposite motives across cultural groups. Last, behavioral convergence starts in our model when environmentalists

reach a lower bound on carbon use. With richer demand systems, in which households reduce their carbon-intensive consumption in proportion to its level, it would start earlier. Such an extension is likely to reinforce our mechanism, since the cultural cost of carbon pricing would then appear at an earlier stage of the transition.

Although our application concerns climate policy, the mechanism is more general. It may arise whenever an intervention changes the incentives attached to an observable behavior, and therefore the extent to which heterogeneous preferences are expressed in actions. Taxes, subsidies, regulations or changes in the available choice set may then induce individuals with different motivations to behave alike. When preferences are transmitted through the behavior they generate, such convergence weakens one of the channels through which these preferences are transmitted across generations. The long-run consequences of a policy cannot then be inferred from its contemporaneous success alone: they also depend on whose behavior adjusts, and on how this adjustment reshapes the cultural constituency on which future policy depends.

Appendix

A. Static Equilibrium

This appendix provides the proofs of Lemmas 1, 2 and 3. In Sections A.1 and A.2, we characterize household behavior and the political equilibrium, which are then used in Section A.3 to establish the properties of equilibrium payoffs and of the behavioral gap.

A.1 Household behavior and proof of Lemma 1

Using the budget constraint (2) to substitute for c_t^i in the utility function (1), and taking R_t and Q_t as given, a type-i household chooses x_t^i to solve

$$\max_{x_t^i \geq \underline{x}} \left\{ -\tau_t x_t^i - \frac{\rho}{2} \left(x_t^i - \alpha_i \right)^2 \right\}. \tag{A.1}$$

The objective is strictly concave in x_t^i, since its second derivative is equal to $-\rho < 0$. Ignoring the lower bound, the first-order condition is

$$-\tau_t - \rho \left(x_t^i - \alpha_i \right) = 0, \tag{A.2}$$

which gives the unconstrained demand $\alpha_i - \tau_t/\rho$. By strict concavity, the objective is increasing in x_t^i below this level and decreasing above. If $\alpha_i - \tau_t/\rho \geq \underline{x}$, the unconstrained demand is therefore feasible and optimal. Otherwise, the objective is decreasing on the whole feasible set, and the household chooses $\underline{x}$. From there, we obtain (5):

$$x_i^*(\tau_t) = \max \left\{ \underline{x},\ \alpha_i - \frac{\tau_t}{\rho} \right\}. \tag{A.3}$$

The lower bound binds as soon as $\alpha_i - \tau_t/\rho \leq \underline{x}$, that is, as soon as $\tau_t \geq \bar{\tau}_i \equiv \rho(\alpha_i - \underline{x})$. Since $\alpha_m > \alpha_e$ by Assumption 1, we have $\bar{\tau}_e < \bar{\tau}_m$. As a result, both types are unconstrained for $\tau_t < \bar{\tau}_e$, only environmentalists are constrained for $\bar{\tau}_e \leq \tau_t < \bar{\tau}_m$, and both types are constrained for $\tau_t \geq \bar{\tau}_m$. A regime in which only materialists are constrained cannot occur, since $\tau_t \geq \bar{\tau}_m$ implies $\tau_t > \bar{\tau}_e$. This

establishes Lemma 1.

A.2 Political equilibrium and proof of Lemma 2

Electoral competition leads both parties to maximize the average utility of the current generation, $\mathcal{W}(\tau_t, q_t)$, given by (16). Since carbon-tax revenues are fully rebated, $R_t = \tau_t X(\tau_t, q_t)$, and tax payments cancel out with the transfer when utilities are aggregated. Using the indirect utility (15), the political objective is then given by (17), (18) and (19) in the three regimes. The proof proceeds in two steps. First, we study the political objective within each regime. Then, we show that the tax obtained in this way is the unique global maximizer.

Within each regime. Differentiating (17) with respect to τ_t gives

$$\frac{\partial \mathcal{W}^{UU}(\tau_t, q_t)}{\partial \tau_t} = \frac{1}{\rho}\Big\{\delta\left[q_t\sigma_e + (1-q_t)\sigma_m\right] - \tau_t\Big\}, \qquad \frac{\partial^2 \mathcal{W}^{UU}(\tau_t, q_t)}{\partial \tau_t^2} = -\frac{1}{\rho} < 0. \tag{A.4}$$

The first term is the marginal environmental benefit of the tax: a higher tax reduces aggregate carbon use by $1/\rho$, which improves environmental quality by δ/ρ, an improvement valued at $q_t\sigma_e + (1-q_t)\sigma_m$ on average. The second term is the marginal adjustment cost borne by households. In the same way, differentiating (18) gives

$$\frac{\partial \mathcal{W}^{BU}(\tau_t, q_t)}{\partial \tau_t} = \frac{1-q_t}{\rho}\Big\{\delta\left[q_t\sigma_e + (1-q_t)\sigma_m\right] - \tau_t\Big\}, \qquad \frac{\partial^2 \mathcal{W}^{BU}(\tau_t, q_t)}{\partial \tau_t^2} = -\frac{1-q_t}{\rho} < 0, \tag{A.5}$$

for $q_t < 1$. In this regime, only materialists adjust to the tax, such that both the marginal environmental benefit and the marginal adjustment cost are multiplied by $1 - q_t$. Finally, $\mathcal{W}^{BB}$ does not depend on τ_t, since carbon use is fixed at $\underline{x}$ and revenues are rebated. It follows that, in both the UU and the BU regimes, the marginal effect of the tax on the political objective has the sign of $\tau^*(q_t) - \tau_t$, where

$$\tau^*(q_t) \equiv \delta\left[q_t\sigma_e + (1-q_t)\sigma_m\right] = \delta\sigma_m + \delta(\sigma_e - \sigma_m)q_t. \tag{A.6}$$

Global maximum. The indirect utilities (15) are continuous at $\bar{\tau}_e$ and $\bar{\tau}_m$, and so is $\mathcal{W}(\tau_t, q_t)$. Under Assumption 2, we have $\tau^*(q_t) \leq \delta\sigma_e < \bar{\tau}_m$ for all $q_t \in [0, 1)$. From the previous paragraph, $\mathcal{W}(\tau_t, q_t)$ is therefore strictly increasing in τ_t for $\tau_t < \tau^*(q_t)$, strictly decreasing for $\tau^*(q_t) < \tau_t < \bar{\tau}_m$, and constant for $\tau_t \geq \bar{\tau}_m$, at a level equal to $\mathcal{W}(\bar{\tau}_m, q_t) < \mathcal{W}(\tau^*(q_t), q_t)$. Consequently, $\tau_t^* = \tau^*(q_t)$ is the unique maximizer of the political objective, and both parties propose it in equilibrium. Note that the last inequality of Assumption 2 is what guarantees uniqueness: without it, any tax above $\bar{\tau}_m$ would maximize the political objective as soon as $\tau^*(q_t) \geq \bar{\tau}_m$.

Regimes along the political-equilibrium path. The equilibrium belongs to the BU regime if and only if $\tau^*(q_t) \geq \bar{\tau}_e$, that is, if and only if

$$\delta\sigma_m + \delta(\sigma_e - \sigma_m)q_t \geq \rho(\alpha_e - \underline{x}).$$

Solving this inequality with respect to q_t gives $q_t \geq \bar{q}$, where $\bar{q}$ is given by (22). The first two inequalities of Assumption 2, $\delta\sigma_m < \rho(\alpha_e - \underline{x}) < \delta\sigma_e$, imply that the numerator of (22) is positive and smaller than

its denominator, such that $0 < \bar{q} < 1$. Finally, since $\tau^*(q_t) < \bar{\tau}_m$ for all $q_t < 1$, materialists are never constrained along the political-equilibrium path. This completes the proof of Lemma 2.

A.3 Relative payoffs, behavioral gap and proof of Lemma 3

Recall that $D(q_t) \equiv V_e(q_t) - V_m(q_t)$, where $V_i(q_t) = v_i(\tau^*(q_t), q_t)$. Since the transfer R_t is common to all households, it cancels out from $D(q_t)$. In what follows, we use

$$\frac{d\tau^*(q_t)}{dq_t} = \delta(\sigma_e - \sigma_m). \tag{A.7}$$

The UU regime. When both types are unconstrained, (15) gives

$$D(q_t) = (\sigma_e - \sigma_m)\left[\bar{Q} - \delta\left(q_t\alpha_e + (1-q_t)\alpha_m - \frac{\tau^*(q_t)}{\rho}\right)\right] + (\alpha_m - \alpha_e)\tau^*(q_t). \tag{A.8}$$

The first term is the additional value that environmentalists attach to environmental quality. The second term reflects their lower exposure to the tax: since they consume $\alpha_m - \alpha_e$ less of the carbon-intensive good, they pay $(\alpha_m - \alpha_e)\tau^*(q_t)$ less in taxes, while both types bear the same adjustment cost. Using (A.7), we obtain

$$\frac{dD(q_t)}{dq_t} = 2\delta(\sigma_e - \sigma_m)(\alpha_m - \alpha_e) + \frac{\delta^2(\sigma_e - \sigma_m)^2}{\rho} > 0, \qquad \frac{d^2D(q_t)}{dq_t^2} = 0. \tag{A.9}$$

The relative payoff of environmentalists is therefore increasing and affine in the UU regime.

The BU regime. When environmentalists are constrained, (15) gives

$$\begin{aligned} D(q_t) = (\sigma_e - \sigma_m)&\left[\bar{Q} - \delta\left(q_t\underline{x} + (1-q_t)\left(\alpha_m - \frac{\tau^*(q_t)}{\rho}\right)\right)\right] \\ &+ (\alpha_m - \underline{x})\tau^*(q_t) - \frac{\tau^*(q_t)^2}{2\rho} - \frac{\rho}{2}(\alpha_e - \underline{x})^2. \end{aligned} \tag{A.10}$$

Differentiating with respect to q_t and using $\Delta x(q_t) = \alpha_m - \underline{x} - \tau^*(q_t)/\rho$, we obtain, after some calculations,

$$\frac{dD(q_t)}{dq_t} = \delta(\sigma_e - \sigma_m)\left[2\Delta x(q_t) + \frac{\delta(\sigma_e - \sigma_m)}{\rho}(1 - q_t)\right] > 0. \tag{A.11}$$

The first term inside the brackets gathers two effects that are proportional to the behavioral gap, which is positive as long as materialists are unconstrained. On the one hand, replacing a materialist by an environmentalist directly reduces aggregate carbon use by $\Delta x(q_t)$, an improvement in environmental quality from which environmentalists benefit relatively more. On the other hand, the higher tax induced by a larger environmentalist share costs environmentalists less than materialists, since they consume $\Delta x(q_t)$ less of the carbon-intensive good. The second term reflects the reduction in the carbon use of materialists induced by this higher tax, which further improves environmental quality. Differentiating once more, we get

$$\frac{d^2D(q_t)}{dq_t^2} = -\frac{3\delta^2(\sigma_e - \sigma_m)^2}{\rho} < 0. \tag{A.12}$$

The relative payoff of environmentalists therefore keeps increasing in the BU regime, but at a decreasing rate. Moreover, $D(q_t)$ is continuous at $\bar{q}$, since $\tau^*(\bar{q}) = \bar{\tau}_e$ implies that the unconstrained demand of environmentalists is exactly equal to $\underline{x}$ at the regime boundary.

The behavioral gap. From (23), $\Delta x(q_t) = \alpha_m - \alpha_e$ for $q_t < \bar{q}$. For $\bar{q} \leq q_t < 1$, using (A.6) and $\tau^*(\bar{q}) = \bar{\tau}_e = \rho(\alpha_e - \underline{x})$, we obtain

$$\Delta x(q_t) = \alpha_m - \alpha_e - \frac{\delta(\sigma_e - \sigma_m)}{\rho}\left(q_t - \bar{q}\right), \tag{A.13}$$

such that $d\Delta x(q_t)/dq_t = -\delta(\sigma_e - \sigma_m)/\rho < 0$. Equations (A.9) to (A.13) establish Lemma 3.

B. Cultural Dynamics

This appendix first derives the adoption rule of children exposed to mixed influences, together with the two variants discussed in the footnotes of Section 3.1. It then proves Proposition 1.

B.1 Trait choice

Baseline rule. A child exposed to mixed influences observes the signal $\tilde{d} = \Delta x_t + \omega\xi$, with $\xi \sim \mathcal{N}(0,1)$. Under the first hypothesis, $\tilde{d} \sim \mathcal{N}(\bar{r}, \omega^2)$, whereas under the second, $\tilde{d} \sim \mathcal{N}(0, \omega^2)$. The ratio of the two likelihoods is

$$\frac{\exp\left[-(\tilde{d} - \bar{r})^2/(2\omega^2)\right]}{\exp\left[-\tilde{d}^2/(2\omega^2)\right]} = \exp\left[\frac{\bar{r}\left(2\tilde{d} - \bar{r}\right)}{2\omega^2}\right]. \tag{B.1}$$

This ratio is at least equal to one if and only if $\tilde{d} \geq \bar{r}/2 = s$, which gives (25). Since $\tilde{d} - \Delta x_t = \omega\xi$ is distributed according to Φ, which is symmetric around zero, the probability of adoption is $\Pr(\omega\xi \geq s - \Delta x_t) = \Phi(\Delta x_t - s)$.

Skeptical prior. Suppose now that the child attaches prior probability $h \in (0,1)$ to the second hypothesis, and adopts the environmental trait when the posterior probability of the first hypothesis is at least one half. Using (B.1), this condition is

$$\frac{\bar{r}\left(2\tilde{d} - \bar{r}\right)}{2\omega^2} \geq \ln\left(\frac{h}{1-h}\right), \qquad \text{that is,} \qquad \tilde{d} \geq \frac{\bar{r}}{2} + \frac{\omega^2}{\bar{r}}\ln\left(\frac{h}{1-h}\right).$$

The threshold s is thus replaced by $\bar{r}/2 + (\omega^2/\bar{r})\ln[h/(1-h)]$, which increases with h.

Sophisticated child. Suppose finally that the child observes the constraints of her role models, such that, under the first hypothesis, she expects the equilibrium gap Δx_t rather than $\bar{r}$. Replacing $\bar{r}$ by Δx_t in the previous condition, she adopts the environmental trait if and only if

$$\tilde{d} \geq \frac{\Delta x_t}{2} + \frac{\omega^2}{\Delta x_t}\ln\left(\frac{h}{1-h}\right).$$

Since $\tilde{d} \sim \mathcal{N}(\Delta x_t, \omega^2)$, the probability of adoption is

$$N\left(\frac{\Delta x_t}{2\omega} - \frac{\omega}{\Delta x_t}\ln\left(\frac{h}{1-h}\right)\right) = N\left(\frac{\Delta x_t^2 - \tilde{s}^2}{2\omega\Delta x_t}\right), \qquad \tilde{s} \equiv \omega\left[2\ln\left(\frac{h}{1-h}\right)\right]^{1/2}, \tag{B.2}$$

where N denotes the cumulative distribution function of the standard normal distribution and $h > 1/2$. The argument of N has derivative $\left(1 + \tilde{s}^2/\Delta x_t^2\right)/(2\omega) > 0$ with respect to Δx_t, such that the probability of adoption increases with the behavioral gap. Moreover, it is below one half if and only if $\Delta x_t < \tilde{s}$. When $h \leq 1/2$, the probability of adoption is always at least one half, and behavioral convergence weakens the transmission of environmental values without reversing it.

B.2 Proof of Proposition 1

Let us first note that G maps $(0,1)$ into itself: for every $q \in (0,1)$, $G(q) = q^2 + 2q(1-q)\Phi(\Delta x(q) - s)$ lies strictly between q^2 and $q^2 + 2q(1-q) = 1 - (1-q)^2$. Moreover, from (28), for every $q_t \in (0,1)$,

$$q_{t+1} \gtrless q_t \quad \Longleftrightarrow \quad \Phi(\Delta x(q_t) - s) \gtrless \frac{1}{2} \quad \Longleftrightarrow \quad \Delta x(q_t) \gtrless s, \tag{B.3}$$

since Φ is strictly increasing and $\Phi(0) = 1/2$. The direction of cultural change is therefore entirely determined by the position of the behavioral gap with respect to s.

Part (i). For $q_t < \bar{q}$, $\Delta x(q_t) = \bar{r} = 2s > s$, since $\alpha_m > \alpha_e$. For $\bar{q} \leq q_t < 1$, the behavioral gap is given by (A.13). It is continuous at $\bar{q}$, where it is equal to $\bar{r} > s$, and strictly decreasing in q_t. As q_t tends to one, it tends to

$$\alpha_m - \underline{x} - \frac{\delta\sigma_e}{\rho},$$

which is lower than $s = (\alpha_m - \alpha_e)/2$ if and only if $\delta\sigma_e > \rho\left[(\alpha_m + \alpha_e)/2 - \underline{x}\right] = \tau^C$, that is, under Assumption 3. There exists therefore a unique $q_H^* \in (\bar{q}, 1)$ such that $\Delta x(q_H^*) = s$. Using (A.13), this condition can be written as

$$\frac{\delta(\sigma_e - \sigma_m)}{\rho}(q_H^* - \bar{q}) = \bar{r} - s = \frac{\alpha_m - \alpha_e}{2},$$

which gives the second expression in (31). Since $\Delta x(q_t) = \alpha_m - \underline{x} - \tau^*(q_t)/\rho$ in the BU regime, the condition $\Delta x(q_H^*) = s$ is also equivalent to $\tau^*(q_H^*) = \tau^C$, which gives the first expression in (31).

Part (ii). From part (i), $\Delta x(q_t) > s$ for $0 < q_t < q_H^*$ and $\Delta x(q_t) < s$ for $q_H^* < q_t < 1$. The result then directly follows from (B.3).

Part (iii). Differentiating (27) with respect to q_t gives

$$G'(q_t) = 2q_t + 2(1-2q_t)\Phi(\Delta x(q_t) - s) - \frac{2q_t(1-q_t)\delta(\sigma_e - \sigma_m)}{\rho}\phi(\Delta x(q_t) - s). \tag{B.4}$$

where the last term is equal to zero in the UU regime, in which $\Delta x(q_t)$ does not depend on q_t. At q_H^*, $\Phi(0) = 1/2$, such that the first two terms sum to one and

$$G'(q_H^*) = 1 - 2q_H^*(1 - q_H^*)\,\phi(0)\,\frac{\delta(\sigma_e - \sigma_m)}{\rho}. \tag{B.5}$$

Since the second term is positive, $G'(q_H^*) < 1$. The steady state is therefore locally stable if and only if $G'(q_H^*) > -1$, which is equivalent to (32). This completes the proof of Proposition 1.

Note finally that $q_t = 0$ and $q_t = 1$ are also steady states of (27), but both are unstable. Indeed, (B.4) gives $G'(0) = 2\Phi(s) > 1$ and, as q_t tends to one, $G'(q_t)$ tends to $2 - 2\Phi\big(\alpha_m - \underline{x} - \delta\sigma_e/\rho - s\big) > 1$ under Assumption 3. From any initial composition $q_0 \in (0,1)$, the share of environmentalists therefore always moves in the direction of q_H^*.

C. Policy Design

This appendix proves Proposition 2 and derives the comparative statics with respect to the lower bound discussed in Section 4.3.

C.1 Proof of Proposition 2

Consider an economy strictly inside the BU regime, with a given cultural composition $q \in (0,1)$. Aggregate carbon use and the behavioral gap are given by (34), such that

$$\frac{\partial X}{\partial \tau} = -\frac{1-q}{\rho}, \qquad \frac{\partial X}{\partial \underline{x}} = q, \qquad \frac{\partial \Delta x}{\partial \tau} = -\frac{1}{\rho}, \qquad \frac{\partial \Delta x}{\partial \underline{x}} = -1. \tag{C.1}$$

Holding q fixed, the share of environmentalists in the next generation, $q^2 + 2q(1-q)\Phi(\Delta x - s)$, responds to a marginal change in the behavioral gap according to

$$dq_{t+1} = 2q(1-q)\,\phi(\Delta x - s)\,d\Delta x. \tag{C.2}$$

Abatement through the carbon tax. From (C.1), reducing aggregate carbon use by $\eta > 0$ through the tax requires $d\tau = \rho\eta/(1-q)$. Since the tax only affects materialists, the behavioral gap falls by $d\Delta x = -d\tau/\rho = -\eta/(1-q)$. Substituting into (C.2) gives

$$dq_{t+1}^{\tau} = 2q(1-q)\,\phi(\Delta x - s)\left(-\frac{\eta}{1-q}\right) = -2q\,\phi(\Delta x - s)\,\eta < 0.$$

Abatement through the lower bound. Reducing aggregate carbon use by the same amount through the lower bound requires $d\underline{x} = -\eta/q$. Since the lower bound only affects environmentalists, the behavioral gap rises by $d\Delta x = -d\underline{x} = \eta/q$. Substituting into (C.2) gives

$$dq_{t+1}^{\underline{x}} = 2q(1-q)\,\phi(\Delta x - s)\,\frac{\eta}{q} = 2(1-q)\,\phi(\Delta x - s)\,\eta > 0.$$

The two interventions generate the same reduction in current carbon use, but the first one is supplied by materialists, which reduces the behavioral gap, whereas the second one is supplied by environmentalists, which increases it. Since $\phi > 0$, they have opposite effects on the share of environmentalists in

the next generation. This proves Proposition 2.

C.2 Carbon dependence and compensation

From (22), (30) and (31),

$$\frac{d\bar{q}}{d\underline{x}} = \frac{dq_H^*}{d\underline{x}} = -\frac{\rho}{\delta(\sigma_e - \sigma_m)} < 0. \tag{C.3}$$

A lower $\underline{x}$ therefore raises both the share at which environmentalists reach the lower bound and the steady-state share of environmentalists. Moreover, Assumption 3 is violated, and the cultural reversal disappears, as soon as $\underline{x} \leq (\alpha_m + \alpha_e)/2 - \delta\sigma_e/\rho$. Environmentalists are never constrained along the political-equilibrium path as soon as $\underline{x} < \alpha_e - \delta\sigma_e/\rho$. Finally, since the demand (A.3) does not depend on R_t, a lump-sum compensation leaves carbon-intensive consumption, the behavioral gap, and therefore $\bar{q}$ and q_H^*, unchanged.

D. Welfare Analysis

This appendix provides the proofs of Lemma 5, Proposition 3, Corollary 1, Proposition 4 and Corollary 2. It also shows that these results hold under a criterion in which the social valuation of environmental quality does not depend on the cultural composition.

D.1 Proof of Lemma 5

In the *BU* regime, environmentalists consume $\underline{x}$, which costs them $\rho(\alpha_e - \underline{x})^2/2 = \bar{\tau}_e^2/(2\rho)$, while materialists consume $\alpha_m - \tau_t/\rho$, which costs them $\tau_t^2/(2\rho)$. Current welfare is then given by (37), with $Q(\tau_t, q_t) = \bar{Q} - \delta\underline{x} - \delta(1 - q_t)\Delta x(\tau_t)$. Holding τ_t fixed, we have $\partial Q(\tau_t, q_t)/\partial q_t = \delta\Delta x(\tau_t)$: replacing one materialist by one environmentalist reduces aggregate carbon use by the behavioral gap. Differentiating (37) with respect to q_t then gives

$$\frac{\partial \mathcal{W}(\tau_t, q_t)}{\partial q_t} = (\sigma_e - \sigma_m)Q(\tau_t, q_t) + \delta\left[q_t\sigma_e + (1 - q_t)\sigma_m\right]\Delta x(\tau_t) + \frac{\tau_t^2 - \bar{\tau}_e^2}{2\rho}. \tag{D.1}$$

Along the political rule, $\mathcal{W}^P(q_t) = \mathcal{W}(\tau^*(q_t), q_t)$, such that

$$\frac{d\mathcal{W}^P(q_t)}{dq_t} = \frac{\partial \mathcal{W}(\tau_t, q_t)}{\partial q_t} + \frac{\partial \mathcal{W}(\tau_t, q_t)}{\partial \tau_t} \times \frac{d\tau^*(q_t)}{dq_t}, \qquad \text{evaluated at } \tau_t = \tau^*(q_t).$$

From Appendix A.2, the second term is equal to zero, since $\tau^*(q_t)$ maximizes current welfare. Using $\delta[q_t\sigma_e + (1 - q_t)\sigma_m] = \tau^*(q_t)$ in (D.1), we obtain (42):

$$\frac{d\mathcal{W}^P(q_t)}{dq_t} = (\sigma_e - \sigma_m)Q^P(q_t) + \tau^*(q_t)\Delta x(q_t) + \frac{\tau^*(q_t)^2 - \bar{\tau}_e^2}{2\rho}. \tag{D.2}$$

In the *BU* regime, $\tau^*(q_t) \geq \bar{\tau}_e$ and $\Delta x(q_t) > 0$, such that the last two terms are respectively non-negative and positive. Whenever $Q^P(q_t) \geq 0$, $d\mathcal{W}^P(q_t)/dq_t > 0$, which proves Lemma 5.

In the *UU* regime, both types bear the same adjustment cost $\tau_t^2/(2\rho)$, and the same steps give $d\mathcal{W}^P(q_t)/dq_t = (\sigma_e - \sigma_m)Q^P(q_t) + \tau^*(q_t)(\alpha_m - \alpha_e)$. Note that the two expressions coincide at $\bar{q}$, where $\tau^*(\bar{q}) = \bar{\tau}_e$ and $\Delta x(\bar{q}) = \alpha_m - \alpha_e$: $\mathcal{W}^P$ is continuously differentiable at the regime boundary.

D.2 Proof of Proposition 3

The proof is performed in three steps. We first compute the effect of the current tax on the cultural composition of the next generation. We then compute the marginal value of the cultural composition at q_H^*. Finally, we combine the two to sign the effect of a reduction of the tax.

Effect of the tax on cultural transmission. From (38), $\partial H(q_t;\tau_t)/\partial\tau_t = 2q_t(1-q_t)\phi(\Delta x(\tau_t) - s)\, d\Delta x(\tau_t)/d\tau_t$. In the UU regime, $\Delta x(\tau_t) = \alpha_m - \alpha_e$ does not depend on the tax, such that $\partial H(q_t;\tau_t)/\partial\tau_t = 0$. In the BU regime, $d\Delta x(\tau_t)/d\tau_t = -1/\rho$, which gives (47).

Value of the cultural composition. By definition, $J^P(q) = \sum_{k\geq 0}\beta^k \mathcal{W}^P(G^k(q))$, where G^k denotes the k-th iterate of G. Under (32), $|G'(q_H^*)| < 1$. Since G' is continuous in the BU regime, there exist $\lambda < 1$ and a neighborhood $\mathcal{N}$ of q_H^* such that $|G'(q)| \leq \lambda$ on $\mathcal{N}$. For every $q \in \mathcal{N}$, we then have $|G(q) - q_H^*| \leq \lambda |q - q_H^*|$, such that G maps $\mathcal{N}$ into itself. On $\mathcal{N}$, the series of derivatives

$$\sum_{k\geq 0}\beta^k \frac{d\mathcal{W}^P(q_k)}{dq}\prod_{j=0}^{k-1} G'(q_j), \qquad \text{with } q_k = G^k(q),$$

is dominated by $\max_{\mathcal{N}}|d\mathcal{W}^P/dq| \times \sum_k(\beta\lambda)^k$, and therefore converges uniformly. J^P is thus differentiable on $\mathcal{N}$, and its derivative is given by this series. At $q = q_H^*$, every q_k is equal to q_H^*, such that

$$\frac{dJ^P(q_H^*)}{dq} = \frac{d\mathcal{W}^P(q_H^*)}{dq}\sum_{k\geq 0}\left[\beta G'(q_H^*)\right]^k = \frac{1}{1-\beta G'(q_H^*)}\frac{d\mathcal{W}^P(q_H^*)}{dq}, \tag{D.3}$$

which is (48). Since $G'(q_H^*) < 1$, the multiplier $1/(1-\beta G'(q_H^*))$ is positive, and $dJ^P(q_H^*)/dq > 0$ whenever $d\mathcal{W}^P(q_H^*)/dq > 0$.

Effect of a reduction of the tax. Since $\bar{\tau}_e < \tau^C < \bar{\tau}_m$, a tax close to τ^C keeps the economy inside the BU regime, and $H(q_H^*;\tau_t)$ belongs to $\mathcal{N}$ for τ_t close enough to τ^C. The function $\Psi(\tau_t; q_H^*) = \mathcal{W}(\tau_t, q_H^*) + \beta J^P(H(q_H^*;\tau_t))$ is therefore differentiable at τ^C. Since $\partial\mathcal{W}(\tau^C, q_H^*)/\partial\tau_t = 0$ and $\Delta x(\tau^C) = s$, (44) and (47) give

$$\left.\frac{\partial\Psi(\tau_t;q_H^*)}{\partial\tau_t}\right|_{\tau_t=\tau^C} = -\Omega(q_H^*) = -\frac{2\beta q_H^*(1-q_H^*)\phi(0)}{\rho}\frac{dJ^P(q_H^*)}{dq} < 0. \tag{D.4}$$

As a result, there exists $\bar{\varepsilon} > 0$ such that $\Psi(\tau^C - \varepsilon; q_H^*) > \Psi(\tau^C; q_H^*)$ for every $\varepsilon \in (0, \bar{\varepsilon})$. This proves Proposition 3.

Orders of magnitude. The two effects of a reduction of the tax by ε can be computed explicitly. Since current welfare (37) is a quadratic function of τ_t, with $\partial^2\mathcal{W}/\partial\tau_t^2 = -(1-q_H^*)/\rho$, and is maximized at τ^C, the current loss is exactly

$$\mathcal{W}(\tau^C, q_H^*) - \mathcal{W}(\tau^C - \varepsilon, q_H^*) = (1-q_H^*)\frac{\varepsilon^2}{2\rho}. \tag{D.5}$$

Moreover, $\Delta x(\tau^C - \varepsilon) - s = \varepsilon/\rho$. Since ϕ is symmetric, $\phi'(0) = 0$, and a Taylor expansion of Φ around zero gives

$$H(q_H^*; \tau^C - \varepsilon) - q_H^* = 2q_H^*(1 - q_H^*)\left[\Phi\left(\frac{\varepsilon}{\rho}\right) - \frac{1}{2}\right] = \frac{2q_H^*(1 - q_H^*)\phi(0)}{\rho}\,\varepsilon + O(\varepsilon^3). \tag{D.6}$$

The current loss is proportional to ε^2, whereas the increase in the share of environmentalists in the next generation is proportional to ε, which gives the approximation (49).

D.3 Proof of Corollary 1

Suppose that G is non-decreasing and $Q^P(q) \geq 0$ on $[\bar{q}, 1)$. The proof is performed in three steps.

Invariance and convergence. Since $G(\bar{q}) > \bar{q}$ and G is non-decreasing, $G(q) \geq G(\bar{q}) > \bar{q}$ for every $q \in [\bar{q}, 1)$. Moreover, $G(q) < 1$ for $q < 1$. The interval $[\bar{q}, 1)$ is therefore invariant under the political dynamics. By Proposition 1, $G(q) > q$ for $q < q_H^*$ and $G(q) < q$ for $q > q_H^*$. Since G is non-decreasing, $G(q) \leq G(q_H^*) = q_H^*$ for $q \leq q_H^*$ and $G(q) \geq q_H^*$ for $q \geq q_H^*$. Every path starting in $(\bar{q}, 1)$ is thus monotone and converges to q_H^*. Note also that $G'(q_H^*) \geq 0$ and, by (B.5), $G'(q_H^*) < 1$, such that q_H^* is locally stable.

Monotonicity of continuation welfare. By Lemma 5, $\mathcal{W}^P$ is strictly increasing on $[\bar{q}, 1)$. For every $k \geq 1$, G^k is non-decreasing, such that $\mathcal{W}^P \circ G^k$ is non-decreasing. J^P, as the sum of a strictly increasing function and of non-decreasing functions, is therefore strictly increasing on $[\bar{q}, 1)$.

Sign of the marginal cultural cost. Since every path converges to q_H^*, it eventually enters the neighborhood $\mathcal{N}$ defined in Appendix D.2, and the series of derivatives converges. J^P is therefore differentiable on $(\bar{q}, 1)$, with

$$\frac{dJ^P(q)}{dq} = \sum_{k \geq 0} \beta^k \frac{d\mathcal{W}^P(q_k)}{dq} \prod_{j=0}^{k-1} G'(q_j) \geq \frac{d\mathcal{W}^P(q)}{dq} > 0, \tag{D.7}$$

since every term of the sum is non-negative. Using (50), we obtain $\Omega(q) > 0$ for every $q \in (\bar{q}, 1)$, and, by (46), $\partial\Psi(\tau_t; q)/\partial\tau_t < 0$ at $\tau_t = \tau^*(q)$. By contrast, $\partial H(q; \tau_t)/\partial\tau_t = 0$ in the UU regime, such that $\Omega(q) = 0$ for $q < \bar{q}$. This proves Corollary 1.

D.4 Proof of Proposition 4

Cultural dynamics under commitment. Consider a constant tax $\widehat{\tau} \in (\bar{\tau}_e, \tau^C)$. Since $\tau^C = \rho(\alpha_m - \underline{x} - s)$, the behavioral gap is

$$\Delta x(\widehat{\tau}) = \alpha_m - \underline{x} - \frac{\widehat{\tau}}{\rho} > s,$$

whatever the cultural composition. From (28), the cultural dynamics is then given by

$$\widehat{q}_{t+1} - \widehat{q}_t = \widehat{q}_t(1 - \widehat{q}_t)\left[2\Phi(\Delta x(\widehat{\tau}) - s) - 1\right], \tag{D.8}$$

where the term in brackets is a positive constant. For any $q_0 \in (0,1)$, the sequence $\{\widehat{q}_t\}$ is therefore increasing and bounded above by one. It converges to a steady state of (D.8) larger than q_0, that is, $\widehat{q}_t \to 1$. Since $\widehat{X}_t = \underline{x} + (1-\widehat{q}_t)\Delta x(\widehat{\tau})$, we also have $\widehat{X}_t \to \underline{x}$.

Long-run welfare. Current welfare is continuous in q_t for a given tax. Using (37) at $q_t = 1$, flow welfare under commitment converges to $\widehat{\mathcal{W}}^{\infty}$, given by (51). Under the political rule, any initial composition in the basin of attraction of q_H^* leads to $q_t^P \to q_H^*$ and $\tau^*(q_t^P) \to \tau^C$. Since $\Delta x(q_H^*) = s$, flow welfare converges to $\mathcal{W}^{P,\infty}$, given by (53). Subtracting the latter from the former and using $\delta[q_H^*\sigma_e + (1-q_H^*)\sigma_m] = \tau^C$ and $\bar{Q} - \delta\underline{x} = Q^{P,\infty} + \delta(1-q_H^*)s$, we obtain (54). Comparing this gap with (D.2) at q_H^*, we can write

$$\widehat{\mathcal{W}}^{\infty} - \mathcal{W}^{P,\infty} = (1-q_H^*)\left[\frac{d\mathcal{W}^P(q_H^*)}{dq} + \delta(\sigma_e - \sigma_m)(1-q_H^*)s\right], \tag{D.9}$$

such that $d\mathcal{W}^P(q_H^*)/dq > 0$ implies $\widehat{\mathcal{W}}^{\infty} > \mathcal{W}^{P,\infty}$.

Discounted welfare. Let $d_t \equiv \widehat{\mathcal{W}}_t - \mathcal{W}_t^P$ denote the difference in flow welfare between the two paths at date t. This sequence does not depend on β, is bounded, and converges to $d_\infty \equiv \widehat{\mathcal{W}}^{\infty} - \mathcal{W}^{P,\infty} > 0$. By the Abelian theorem for power series,

$$\lim_{\beta \to 1} (1-\beta) \sum_{t=0}^{\infty} \beta^t d_t = d_\infty > 0. \tag{D.10}$$

There exists therefore $\bar{\beta}(q_0, \widehat{\tau}) < 1$ such that $\widehat{\mathcal{V}}(q_0) - \mathcal{V}^P(q_0) = \sum_t \beta^t d_t > 0$ for every $\beta > \bar{\beta}(q_0, \widehat{\tau})$. Finally, $\widehat{X}^{\infty} = \underline{x} < \underline{x} + (1-q_H^*)s = X^{P,\infty}$. This proves Proposition 4.

D.5 Environmental and fixed-reference rankings

Proof of Corollary 2. From the previous subsection, environmental quality converges to $\widehat{Q}^{\infty} = \bar{Q} - \delta\underline{x}$ under commitment, and to $Q^{P,\infty} = \bar{Q} - \delta\left[\underline{x} + (1-q_H^*)s\right]$ under the political rule, which gives (56). Let $e_t \equiv \widehat{Q}_t - Q_t^P$. This sequence is bounded and converges to $\delta(1-q_H^*)s > 0$. The same Abelian argument as above implies that there exists $\bar{\beta}_E(q_0, \widehat{\tau}) < 1$ such that $\widehat{\mathcal{E}}(q_0) > \mathcal{E}^P(q_0)$ for every $\beta > \bar{\beta}_E(q_0, \widehat{\tau})$. Note that this argument does not use any condition on $d\mathcal{W}^P(q_H^*)/dq$.

Fixed-reference welfare. To separate our welfare results from changes in the weight that the population attaches to environmental quality, we now fix the social valuation of environmental quality at its steady-state level, $\sigma^R \equiv q_H^*\sigma_e + (1-q_H^*)\sigma_m$, such that $\delta\sigma^R = \tau^C$. In the *BU* regime, current welfare becomes

$$\mathcal{W}^R(\tau_t, q_t) = w + \sigma^R Q(\tau_t, q_t) - q_t\frac{\bar{\tau}_e^2}{2\rho} - (1-q_t)\frac{\tau_t^2}{2\rho}. \tag{D.11}$$

This criterion takes into account the actual adjustment costs of households, but values environmental quality in the same way whatever the cultural composition. Differentiating with respect to τ_t, we get $\partial\mathcal{W}^R(\tau_t, q_t)/\partial\tau_t = (1-q_t)(\tau^C - \tau_t)/\rho$, which is equal to zero at $\tau_t = \tau^C$. Holding the tax fixed and

differentiating with respect to q_t, we get

$$\frac{\partial \mathcal{W}^R(\tau_t, q_t)}{\partial q_t} = \tau^C \Delta x(\tau_t) + \frac{\tau_t^2 - \bar{\tau}_e^2}{2\rho}. \tag{D.12}$$

Along the political rule, current fixed-reference welfare is $\mathcal{W}^{R,P}(q_t) \equiv \mathcal{W}^R(\tau^*(q_t), q_t)$. At q_H^*, the effect of q_t that goes through the tax is equal to zero, since $\tau^*(q_H^*) = \tau^C$ maximizes $\mathcal{W}^R(\tau_t, q_H^*)$. As a result,

$$\frac{d\mathcal{W}^{R,P}(q_H^*)}{dq} = \tau^C s + \frac{(\tau^C)^2 - \bar{\tau}_e^2}{2\rho} > 0, \tag{D.13}$$

since $s > 0$ and $\tau^C > \bar{\tau}_e$. This expression corresponds to (42) at q_H^* without its first term $(\sigma_e - \sigma_m)Q^P(q_H^*)$, which comes from the change in the population weight placed on environmental quality.

The proof of Proposition 3 can then be repeated with $\mathcal{W}^R$ in place of $\mathcal{W}$. At q_H^*, the political tax maximizes current fixed-reference welfare, continuation welfare increases with the cultural composition by (D.13) and (D.3), and $\partial H(q_H^*; \tau_t)/\partial \tau_t < 0$ at $\tau_t = \tau^C$. A sufficiently small reduction of the current tax therefore also raises discounted fixed-reference welfare. Finally, the long-run gap between the two paths becomes

$$\widehat{\mathcal{W}}^{R,\infty} - \mathcal{W}^{R,P,\infty} = \sigma^R \delta (1 - q_H^*) s + (1 - q_H^*) \frac{(\tau^C)^2 - \bar{\tau}_e^2}{2\rho} = (1 - q_H^*) \left[\tau^C s + \frac{(\tau^C)^2 - \bar{\tau}_e^2}{2\rho} \right] > 0. \tag{D.14}$$

The long-run ranking of Proposition 4, and therefore the ranking of discounted welfare for β close enough to one, survive when environmental quality is valued in the same way under all cultural compositions.

E. Purposeful Parental Socialization

This appendix provides the derivations of Section 6. In Sections E.1 and E.2, we derive the gain that parents attach to an environmentalist child and solve their socialization problem. These results are then used in Section E.3 to prove Proposition 5. Throughout, Assumption 2 holds, such that $\tau_t^* \leq \delta \sigma_e < \bar{\tau}_m$: materialists are never constrained, and $x_t^m = \alpha_m - \tau_t^*/\rho$ for all $q_t < 1$.

E.1 Parental values

From (58), the value that a parent of type i attaches to a child of type j is

$$V_{ij}(q_t) = w + R_t - \tau_t^* x_t^j - \frac{\rho}{2} \left(x_t^j - \alpha_i \right)^2 + \sigma_i Q_t.$$

Since the cultural trait of an individual child affects neither R_t, nor Q_t, nor the political equilibrium, these terms cancel out when the parent compares the two possible traits of her child, and

$$V_{ie}(q_t) - V_{im}(q_t) = -\tau_t^* \left(x_t^e - x_t^m \right) - \frac{\rho}{2} \left[(x_t^e - \alpha_i)^2 - (x_t^m - \alpha_i)^2 \right]. \tag{E.1}$$

For any x, we have

$$-\tau_t^* x - \frac{\rho}{2} (x - \alpha_i)^2 = \frac{(\tau_t^*)^2}{2\rho} - \alpha_i \tau_t^* - \frac{\rho}{2} \left(x - \alpha_i + \frac{\tau_t^*}{\rho} \right)^2.$$

The first two terms on the right-hand side do not depend on x and cancel out in (E.1), which gives (59).

Environmentalist parents. For $i = e$, we have $x_t^m - \alpha_e + \tau_t^*/\rho = \alpha_m - \alpha_e = \bar{r}$ in both regimes. In the UU regime, $x_t^e = \alpha_e - \tau_t^*/\rho$, such that $x_t^e - \alpha_e + \tau_t^*/\rho = 0$. In the BU regime, $x_t^e = \underline{x}$ and $\Delta x(q_t) = \alpha_m - \underline{x} - \tau_t^*/\rho$, such that

$$x_t^e - \alpha_e + \frac{\tau_t^*}{\rho} = \underline{x} - \alpha_e + \frac{\tau_t^*}{\rho} = \bar{r} - \Delta x(q_t) = \frac{\tau_t^* - \bar{\tau}_e}{\rho}.$$

Since $\Delta x(q_t) = \bar{r}$ in the UU regime, both cases are summarized by (60):

$$V_{ee}(q_t) - V_{em}(q_t) = \frac{\rho}{2}\left\{\bar{r}^2 - [\bar{r} - \Delta x(q_t)]^2\right\} = \rho \bar{r}\, \Delta x(q_t) - \frac{\rho}{2}\left[\Delta x(q_t)\right]^2.$$

Using $\tau_t^* - \bar{\tau}_e = \delta(\sigma_e - \sigma_m)(q_t - \bar{q})$ in the BU regime, we obtain (61). This gain is strictly positive for $q_t < 1$, since $0 < \Delta x(q_t) \leq \bar{r}$ under Assumption 2.

Materialist parents. For $i = m$, we have $x_t^m - \alpha_m + \tau_t^*/\rho = 0$ and $x_t^e - \alpha_m + \tau_t^*/\rho = x_t^e - x_t^m = -\Delta x(q_t)$ in both regimes, which gives (62).

E.2 Optimal socialization and proof of Lemma 6

The derivative of the objective (63) with respect to a_t^i is

$$\gamma\left[V_{ie}(q_t) - V_{im}(q_t)\right] - \kappa a_t^i,$$

and its second derivative is equal to $-\kappa < 0$. For environmentalist parents, $V_{ee}(q_t) - V_{em}(q_t) > 0$, such that the unique solution is interior and given by $a_t^{e*} = (\gamma/\kappa)[V_{ee}(q_t) - V_{em}(q_t)] > 0$. Using (61), we obtain (65). This effort is constant for $q_t < \bar{q}$ and, for $\bar{q} < q_t < 1$,

$$\frac{da_t^{e*}}{dq_t} = -\frac{\gamma\delta^2(\sigma_e - \sigma_m)^2}{\kappa\rho}(q_t - \bar{q}) < 0. \tag{E.2}$$

For materialist parents, $V_{me}(q_t) - V_{mm}(q_t) < 0$, such that the objective is strictly decreasing at $a_t^m = 0$ and, by concavity, on the whole feasible set. The non-negativity constraint therefore binds, and $a_t^{m*} = 0$. This proves Lemma 6.

E.3 Proof of Proposition 5

From (66) and Lemma 6, $p_t^{ee} = \mu + \gamma a_t^{e*}$ and $p_t^{me} = \mu$. Since $\gamma a_t^{e*} \leq \gamma^2\rho\bar{r}^2/(2\kappa) = \theta$, the first condition of Assumption 4 ensures that $p_t^{ee} < 1$ for all q_t. Aggregating across parental types gives (68). Let

$$F(q) \equiv \Gamma(q) - q = \begin{cases} \mu - (1-\theta)q, & q < \bar{q}, \\ \mu - (1-\theta)q - \dfrac{\gamma^2\delta^2(\sigma_e - \sigma_m)^2}{2\kappa\rho}\, q\,(q - \bar{q})^2, & \bar{q} \leq q < 1, \end{cases} \tag{E.3}$$

which is continuous on $[0, 1)$.

Part (i). Since $1-\theta > 0$, F is strictly decreasing on $[0, \bar{q})$, and $F(q) > F(\bar{q}) = \mu - (1-\theta)\bar{q} > 0$ for $q < \bar{q}$, by the second condition of Assumption 4. As q tends to one,

$$\lim_{q \to 1} F(q) = \mu + \theta - 1 - \frac{\gamma^2 \delta^2 (\sigma_e - \sigma_m)^2}{2\kappa\rho} (1-\bar{q})^2 < 0,$$

by the first condition of Assumption 4. Moreover, for $\bar{q} \leq q < 1$,

$$\frac{dF(q)}{dq} = -(1-\theta) - \frac{\gamma^2 \delta^2 (\sigma_e - \sigma_m)^2}{2\kappa\rho} (q-\bar{q})(3q-\bar{q}) < 0. \tag{E.4}$$

F is therefore continuous and strictly decreasing on $[\bar{q}, 1)$, with $F(\bar{q}) > 0$ and a negative limit as q tends to one. There exists a unique $q^*_{PS} \in (\bar{q}, 1)$ such that $F(q^*_{PS}) = 0$. Furthermore, $F(q) > 0$, that is $q_{t+1} > q_t$, for $0 \leq q_t < q^*_{PS}$, and $F(q) < 0$, that is $q_{t+1} < q_t$, for $q^*_{PS} < q_t < 1$.

Part (ii). When the behavioral gap is fixed to $\bar{r}$, the mapping is $\Gamma^{\bar{r}}(q) = \mu + \theta q$, whose unique steady state is $q^{\bar{r}}_{PS} = \mu/(1-\theta)$. Assumption 4 implies $q^{\bar{r}}_{PS} < 1$, since $\mu < 1-\theta$, and $q^{\bar{r}}_{PS} > \bar{q}$, since $\mu > (1-\theta)\bar{q}$. Evaluating F at this point, the first two terms cancel out, and

$$F\left(q^{\bar{r}}_{PS}\right) = -\frac{\gamma^2 \delta^2 (\sigma_e - \sigma_m)^2}{2\kappa\rho} q^{\bar{r}}_{PS} \left(q^{\bar{r}}_{PS} - \bar{q}\right)^2 < 0.$$

Since $F(q^*_{PS}) = 0$ and F is strictly decreasing on $[\bar{q}, 1)$, we obtain $q^*_{PS} < q^{\bar{r}}_{PS}$.

Part (iii). Since $\Gamma(q) = F(q) + q$, (E.4) gives, at $q^*_{PS} > \bar{q}$,

$$\Gamma'(q^*_{PS}) = \theta - \frac{\gamma^2 \delta^2 (\sigma_e - \sigma_m)^2}{2\kappa\rho} (q^*_{PS} - \bar{q})(3q^*_{PS} - \bar{q}).$$

Since $\theta < 1$ and the second term is positive, $\Gamma'(q^*_{PS}) < 1$. The steady state is therefore locally stable if and only if $\Gamma'(q^*_{PS}) > -1$, that is, if and only if (72) holds. This completes the proof of Proposition 5.

References

Abdelgadir, A. and Fouka, V. (2020). Political secularism and muslim integration in the west: Assessing the effects of the french headscarf ban. *American Political Science Review*, 114(3):707–723.

Abnett, K. (2026). Ten countries urge EU to rethink new carbon price on fuel. Reuters. 15 July 2026.

Acemoglu, D., Aghion, P., Bursztyn, L., and Hemous, D. (2012). The environment and directed technical change. *American Economic Review*, 102(1):131–166.

Aghion, P., Dechezleprêtre, A., Hémous, D., Martin, R., and Van Reenen, J. (2016). Carbon taxes, path dependency, and directed technical change: Evidence from the auto industry. *Journal of Political Economy*, 124(1):1–51.

Andersson, J. J. (2019). Carbon taxes and CO2 emissions: Sweden as a case study. *American Economic Journal: Economic Policy*, 11(4):1–30.

Ariely, D., Bracha, A., and Meier, S. (2009). Doing good or doing well? image motivation and monetary incentives in behaving prosocially. *American Economic Review*, 99(1):544–555.

Ascari, G., Colciago, A., Haber, T., and Wöhrmüller, S. (2026). Inequality along the european green transition. *The Economic Journal*, 136(677):1729–1760.

Attari, S. Z., Krantz, D. H., and Weber, E. U. (2016). Statements about climate researchers' carbon footprints affect their credibility and the impact of their advice. *Climatic Change*, 138(1):325–338.

Attari, S. Z., Krantz, D. H., and Weber, E. U. (2019). Climate change communicators' carbon footprints affect their audience's policy support. *Climatic Change*, 154(3):529–545.

Barrage, L. (2020). Optimal dynamic carbon taxes in a climate–economy model with distortionary fiscal policy. *The Review of Economic Studies*, 87(1):1–39.

Baudin, T. and Hiller, V. (2019). On the dynamics of gender differences in preferences. *Oxford Economic Papers*, 71(3):503–527.

Bazzi, S., Hilmy, M., and Marx, B. (2026). Religion, education, and the state. *Review of Economic Studies*, 93(3):1494–1535.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Besley, T. (2017). Aspirations and the Political Economy of Inequality. *Oxford Economic Papers*, 69(1):1–35.

Besley, T. and Ghatak, M. (2016). Market incentives and the evolution of intrinsic motivation. Mimeo.

Besley, T. and Persson, T. (2019a). Democratic values and institutions. *American Economic Review: Insights*, 1(1):59–76.

Besley, T. and Persson, T. (2019b). The dynamics of environmental politics and values. *Journal of the European Economic Association*, 17(4):993–1024.

Besley, T. and Persson, T. (2023). The political economics of green transitions. *The Quarterly Journal of Economics*, 138(3):1863–1906.

Bezin, E. (2015). A cultural model of private provision and the environment. *Journal of Environmental Economics and Management*, 71:109–124.

Bezin, E. (2019). The economics of green consumption, cultural transmission and sustainable technological change. *Journal of Economic Theory*, 181:497–546.

Bigio, S., Känzig, D. R., Sánchez, P., and Walsh, C. (2026). Carbon pricing and inequality: A normative perspective. *The Economic Journal*, 136(677):1659–1682.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319.

Bisin, A. and Verdier, T. (2011). The economics of cultural transmission and socialization. In Benhabib, J., Bisin, A., and Jackson, M. O., editors, *Handbook of Social Economics*, volume 1A, pages 339–416. North-Holland, Amsterdam.

Bisin, A. and Verdier, T. (2024). On the joint evolution of culture and political institutions: Elites and civil society. *Journal of Political Economy*, 132(5):1485–1564.

Boyd, R. and Richerson, P. J. (1985). *Culture and the Evolutionary Process*. University of Chicago Press, Chicago.

Carvalho, J.-P. (2013). Veiling. *The Quarterly Journal of Economics*, 128(1):337–370.

Carvalho, J.-P., Koyama, M., and Williams, C. (2024). Resisting education. *Journal of the European Economic Association*, 22(6):2549–2597.

Conte, B., Desmet, K., and Rossi-Hansberg, E. (2026). On the geographic implications of carbon taxes. *The Economic Journal*, 136(677):1547–1581.

Douenne, T. and Fabre, A. (2022). Yellow vests, pessimistic beliefs, and carbon tax aversion. *American Economic Journal: Economic Policy*, 14(1):81–110.

Douenne, T., Hummel, A. J., and Pedroni, M. (2026). Optimal fiscal policy in a climate-economy model with heterogeneous households. *The Economic Journal*, 136(677):1761–1801.

Eisner, A., Kulmer, V., and Kortschak, D. (2021). Distributional effects of carbon pricing when considering household heterogeneity: An easi application for austria. *Energy Policy*, 156:112478.

Fernández, R. (2013). Cultural change as learning: The evolution of female labor force participation over a century. *American Economic Review*, 103(1):472–500.

Fogli, A. and Veldkamp, L. (2011). Nature or nurture? Learning and the geography of female labor force participation. *Econometrica*, 79(4):1103–1138.

Fouka, V. (2019). How do immigrants respond to discrimination? the case of germans in the US during world war I. *American Political Science Review*, 113(2):405–422.

Fouka, V. (2020). Backlash: The unintended effects of language prohibition in US schools after world war I. *Review of Economic Studies*, 87(1):204–239.

Gilbert, D. T. and Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin*, 117(1):21–38.

Gillingham, K. and Munk-Nielsen, A. (2019). A tale of two tails: Commuting and the fuel price response in driving. *Journal of Urban Economics*, 109:27–40.

Goette, L. and Stutzer, A. (2020). Blood donations and incentives: Evidence from a field experiment. *Journal of Economic Behavior & Organization*, 170:52–74.

Golosov, M., Hassler, J., Krusell, P., and Tsyvinski, A. (2014). Optimal taxes on fossil fuel in general equilibrium. *Econometrica*, 82(1):41–88.

Goulder, L. H., Hafstead, M. A. C., Kim, G., and Long, X. (2019). Impacts of a carbon tax across US household income groups: What are the equity-efficiency trade-offs? *Journal of Public Economics*, 175:44–64.

Grønhøj, A. and Thøgersen, J. (2009). Like father, like son? intergenerational transmission of values, attitudes, and behaviours in the environmental domain. *Journal of Environmental Psychology*, 29(4):414–421.

Grønhøj, A. and Thøgersen, J. (2012). Action speaks louder than words: The effect of personal attitudes and family norms on adolescents' pro-environmental behaviour. *Journal of Economic Psychology*, 33(1):292–302.

Henrich, J. (2009). The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and Human Behavior*, 30(4):244–260.

Herweg, F. and Schmidt, K. M. (2022). How to regulate carbon emissions with climate-conscious consumers. *The Economic Journal*, 132(648):2992–3019.

Hochmuth, P., Krusell, P., and Mitman, K. (2026). Distributional consequences of becoming climate-neutral. *The Economic Journal*, 136(677):1704–1728.

Jia, F. and Yu, H. (2021). Action, communication, and engagement: How parents "ace" children's pro-environmental behaviors. *Journal of Environmental Psychology*, 74:101575.

Jones, E. E. and Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3(1):1–24.

Klenert, D., Mattauch, L., Combet, E., Edenhofer, O., Hepburn, C., Rafaty, R., and Stern, N. (2018). Making carbon pricing work for citizens. *Nature Climate Change*, 8:669–677.

Konradt, M. and Mangiante, G. (2026). The unequal costs of carbon pricing: Economic and political effects across european regions. *The Economic Journal*, 136(677):1683–1703.

Kraft-Todd, G. T., Bollinger, B., Gillingham, K., Lamp, S., and Rand, D. G. (2018). Credibility-enhancing displays promote the provision of non-normative public goods. *Nature*, 563(7730):245–248.

Lanman, J. A. and Buhrmester, M. D. (2017). Religious actions speak louder than words: Exposure to credibility-enhancing displays predicts theism. *Religion, Brain & Behavior*, 7(1):3–16.

Lindbeck, A. and Weibull, J. W. (1987). Balanced-budget redistribution as the outcome of political competition. *Public Choice*, 52(3):273–297.

Ling, M. and Xu, L. (2021). How and when financial incentives crowd out pro-environmental motivation: A longitudinal quasi-experimental study. *Journal of Environmental Psychology*, 78:101715.

Mattauch, L., Hepburn, C., Spuler, F., and Stern, N. (2022). The economics of climate change with endogenous preferences. *Resource and Energy Economics*, 69:101312.

Persson, T. and Tabellini, G. (2000). *Political Economics: Explaining Economic Policy*. MIT Press, Cambridge, MA.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In Berkowitz, L., editor, *Advances in Experimental Social Psychology*, volume 10, pages 173–220. Academic Press.

Squicciarini, M. P. (2020). Devotion and development: Religiosity, education, and economic progress in nineteenth-century france. *American Economic Review*, 110(11):3454–3491.

Timilsina, G. R. (2022). Carbon taxes. *Journal of Economic Literature*, 60(4):1456–1502.

Torren-Peraire, D., van den Bergh, J. C. J. M., and Savin, I. (2025). The cultural multiplier of climate policy. *Environmental and Resource Economics*, 88:3215–3251.

Touré, N. (2021). Culture, institutions and the industrialization process. *Journal of Economic Behavior & Organization*, 186:481–503.

World Bank (2026). State and trends of carbon pricing 2026. Technical report, World Bank, Washington, DC. Press release of 19 May 2026: https://www.worldbank.org/en/news/press-release/2026/05/19/direct-carbon-pricing-covers-nearly-one-third-of-global-emissions.

Xu, L., Yang, H., and Ling, M. (2023). Factors moderating the effect of financial incentives on household recycling: A motivation crowding perspective. *Environmental Impact Assessment Review*, 100:107078.