Autonomy and the Endogenous Formation of Social Norms: Theory and Evidence from Universalism

Laure Athias*

July 12, 2025

Abstract

I provide theory and evidence on *autonomy* – the self-institution of social norms. Rather than emerging from the aggregation of fixed private social preferences, social norms are endogenously created through costly collective normative effort, which rational individuals exert because the norm legitimizes institutional functionality. Autonomy combines: i) individuals may deliberatively diverge from the instituted norm; ii) agency within individuals is layered, comprising a private normative orientation – what one endorses for oneself – and a public normative orientation – what one endorses for society – which may (strategically) diverge; iii) individuals engage in ongoing self-reflection on their support for the instituted norm, shaped by both institutional and peer normative embodiment within a given environment. The model generates propositions regarding optimal normative effort, the formation of steady-state preferences, and endogenous norm erosion and shift. Using individual-level, geolocated panel data from Switzerland, I exploit within-individual variation in the normative composition of state and peer environments to investigate whether universalism exemplifies the dynamics of a self-instituted social norm, combining multiple identification strategies. The results strongly support the model's core autonomy-based assumptions and predictions, underscoring the persistence of the Western European Enlightenment tradition of rationality, critical reflection, and the imperative that all aspects of social life – including moral order – be justified without recourse to divine or natural authority. Overall, the paper identifies the endogenous conditions under which universalism may erode – but also persist – as well as under which democratic participation remains vibrant.

Keywords: Social norms; Social preferences; Autonomy; Institutional symbolism; Institutional functionality; Public sphere; Enlightenment; Universalism; State legitimacy; State capacity; Bureaucracy; Democracy.

"I call autonomous a society that not only knows explicitly that it has created its own laws but has instituted itself so as to free its radical imaginary and enable itself to alter its institutions through collective, self-reflective, and deliberate activity."

Castoriadis (1997), p. 132

"Subjectivity, as agent of reflection and deliberation (as thought and will) is a social-historical project; its origins, repeated twice with different modalities in Greece and in Western Europe, can be dated and located."

Castoriadis (1991), p. 144

1   Introduction

Universalism – the principle that all individuals, regardless of identity, are entitled to equal rights – is often regarded as a foundational social norm in modern Western societies (Taylor (2004), Henrich (2020)). This stands in contrast to context-dependent moral systems, in which obligations are stronger toward kin or in-groups. Universalism has attracted considerable interest because it determines the well-functioning of institutions and economic development (Tabellini (2008b), Henrich (2020)). Yet history reveals the fragility of this normative ideal: nationalist and exclusionary ideologies have repeatedly garnered broad support. A prominent explanation in economics for shifts in social norms is that individuals harbor in-group–biased preferences that remain latent under strong prevailing norms, but are revealed when those normative constraints are exogenously weakened – for example, following new public information about aggregated private preferences (Bursztyn et al. (2020)). While compelling, this perspective implies that universalism is not genuinely internalized but rather upheld by cultural social expectations[1]. This raises a central tension: how can universalism function as a social norm if it lacks genuine internalization? Put differently, if individuals merely conform to universalism under social pressure, what sustains its legitimacy as a shared norm? It also assumes that norm weakening is primarily driven by exogenous shocks, such as election outcomes. However, prominent studies highlight the important role of interpersonal interactions and 'word of mouth' in shaping voting behavior (e.g. Lazarsfeld and Gaudet (1944), Bond et al. (2012), Jones et al. (2017)). These puzzles call for a more rigorous account of the origins and formation of social norms.

In this paper, I build on Castoriadis (1987)'s concept of *autonomy* – the conscious self-institution of norms (autos nomos) – to develop a utility-based model of social norm formation, in which norms are endogenously created through costly normative effort, rather than passively derived from the aggregation of fixed preferences[2]. Individuals derive utility from the formation of a social norm insofar as it legitimizes institutional functionality, thereby fostering voluntary compliance and, in turn, institutional effectiveness (Acemoglu and Jackson (2017), Levi (1997), Tyler (2006)). Thus, the norm is not merely a constraint on behavior but a constitutive principle of institutions. However, utility increases with

---

[1]Henrich (2020) attributes the universal morality of WEIRD (Western, Educated, Industrialized, Rich and Democratic) societies to cultural changes, particularly those driven by the Western Church, that promoted greater psychological independence and impersonal prosociality.

[2]Autonomy implies that social norms are *instituted* – not simply *chosen* – at the societal level. Even when norms appear to reflect "societal preferences", if they emerge through aggregation rules (e.g., Rawls (1971)'s maximin or Harsanyi (1955)'s average utility), the process remains heteronomous: individuals obey a logic external to themselves. In contrast, autonomy means that individuals see themselves not merely as preference holders but as co-creators of social norms.

alignment between the instituted norm and the individual's private normative orientation. This reflects the core of autonomy as theorized by Castoriadis (1987): the capacity to question and evaluate inherited meanings in light of one's own evolving normative orientation, rather than to passively accept them. Furthermore, my model allows for *autonomy-based divergences* between an individual's normative orientation in the private and public realms through a *layered conception of agency*: autonomous individuals can transcend their private preferences and hold a distinct normative orientation for society[3]. Crucially, autonomy implies that individuals' social preferences are not fixed: even within a stable institutional setting, they evolve through an ongoing process of self-reflection shaped by the normative embodiment of both institutions and peers[4]. The model yields four core propositions. First, alignment between an individual's private social preference and the instituted social norm increases their willingness and incentives to exert normative effort. Second, the equilibrium structure predicts both normative convergence and pluralism as endogenous outcomes, depending on the relative strength of institutional and peer normative embodiment. Third, when institutional normative embodiment is weak relative to peer, peer dynamics can endogenously drive the disintegration of established norms through decentralized feedback mechanisms, reflecting the emergence of a *radical imaginary* (Castoriadis (1987)) that reshapes the symbolic order from within. Fourth, when both forces are weak, the system becomes susceptible to norm drift, instability, or persistent fluctuations around equilibrium.

The model applies specifically to autonomous societies. As the opening quote suggests, Castoriadis (1991) identifies only two historical moments when the project of autonomy was explicitly instituted: classical Athens and post-Enlightenment Western Europe. In his view, the Enlightenment's emphasis on rationality, critical reflection, and the idea that all aspects of social life (morals, laws, institutions) are open to questioning – and must be justified without recourse to divine or natural authority – marked a historical rupture[5]. Importantly, autonomy does not imply a specific normative content: universalism is not the necessary outcome of autonomy. Yet in Western Europe, universalism emerged as the dominant norm, rooted in seventeenth-century Natural Law theories. In response to the Wars of Religion, thinkers like Grotius and Locke developed concepts of natural rights

---

[3]Athias and Ventelou (2025) provide empirical evidence of within-individual normative divergences that vary systematically with socio-demographic characteristics. Specifically, they find that individuals with higher cognitive skills are more likely to endorse universalism as a private preference than as a social norm, whereas men and right-leaning individuals exhibit the opposite pattern, endorsing universalism more likely as a social norm than as a private preference.

[4]The influence of institutional normative embodiment is a corollary of the social norm legitimizing institutions. It resonates with Castoriadis (1987)'s dual dimension of institutions: a functional and a symbolic dimension, with the latter shaping individuals' values. See also Bénabou and Tirole (2025) on the expressive role of law as embodying society's values. Thus, institutions are not neutral functional vehicles but norm-bearing entities. The normative embodiment of peers may encompass what Habermas (1989) calls the public sphere: a social space where individuals come together to deliberate on matters of common concern – a space between the private realm of the household and the state – and where public opinion is formed.

[5]Importantly, a society may have elections, rights, and a written constitution and still not be autonomous if it fails to recognize itself as the source of its social meaning, subject to collective questioning and reinvention. As Castoriadis (1996) argues, the United States, despite its rupture with monarchy and colonial rule, did not fully undertake the project of self-institution. Instead, the American Revolution substituted one form of heteronomy for another: it preserved the underlying imaginaries of economic expansion and religious authority. Political liberalism became entwined with a naturalized market order, treating its institutions as self-evident. See also Bénabou et al. (2021) on the American religious equilibrium.

and mutual obligations grounded in a shared humanity rather than in particular identities (Taylor (2004))[6]. My model predicts that universalism achieved normative stability in this context through sustained institutional embodiment.

To empirically test the model and explore the dynamics of universalism as a self-instituted norm, I focus on Switzerland – a country situated within post-Enlightenment Western Europe, whose semi-direct democratic institutions, combined with a tradition of pragmatic consensus, neutrality, and institutional stability, make it an instructive case for assessing the extent of autonomy. I use data from the Swiss Household Panel Survey (SHPS), a panel survey that annually interviews all members of approximately 5,000 randomly selected households, covering over 12,000 individuals since 1999. This dataset includes a precise measure of individual support for universalism as a social norm, captured by the question: "Are you in favour of Switzerland offering foreigners the same opportunities as those offered to Swiss citizens, or in favour of Switzerland offering Swiss citizens better opportunities?". The SHPS is geolocated at the municipality level, and includes occupational information, including whether respondents work for the state. This allows me to construct time-varying measures of the normative composition of individuals' deliberative environments. Specifically, I focus on non-workers (youth over age 18, the unemployed, and retirees) and compute the local shares of individuals – among both state agents and peers – who endorse universalism as a social norm. This strategy allows me to exploit within-individual variation in the normative composition of institutional and social environments and to assess their relative importance in shaping self-deliberation over the social norm. I proceed in four steps. First, I examine whether non workers' universalism (only) evolves in response to changes in the normative composition of their deliberative environment, and compare the relative importance of both deliberative spheres. Second, building upon Enke et al. (2023), I explore heterogeneity by political ideology to assess whether individuals who hold universalism as a personal value remain susceptible to deliberative influence. This would reveal within-individual divergence in normative orientations across layers of agency, consistent with deliberative endorsement rather than coercion or social pressure. Third, I test whether alignment with the prevailing social norm increases democratic participation and collective engagement. Fourth, I assess whether the influence of state agents stems from their institutional role in embodying social norms, by examining preference dynamics among (low-level) bureaucrats who switch between private and public sectors, and by leveraging cultural variation in the social contract underpinning the modern state. Overall, the results strongly support the model's core autonomy-based framework, confirming the hypothesis that universalism is a self-instituted norm, emerging from ongoing deliberative effort shaped by both institutional and peer normative embodiment – underscoring its inherent fragility but also its endogenous sustainability.

This work contributes to both the literature on social norms and the literature on the interplay between culture and institutions by formalizing – and empirically identifying – social norms as endogenous, deliberative social constructs embodied in institutions. Building on Castoriadis (1987)'s philosophical concept of *autonomy* – understood as the capacity to self-institute social norms without recourse to tradition, divine authority, supernatural forces, or economic determinism –, this paper challenges the prevailing view in economics that norms are merely external constraints on behavior, enforced through social image concerns or internal dissonance costs, and that norm change results from be-

---

[6]As Taylor (2004) notes (p. 24): "It often happens that what start off as theories held by a few people come to infiltrate the social imaginary, first of elites, perhaps, and then of the whole society."

lief updating within Bayesian equilibria of social interactions (Bénabou and Tirole (2003), Bénabou and Tirole (2011), Acemoglu and Jackson (2015), Bursztyn et al. (2020), Tirole (2023)). I conceptualize autonomy instead as the emergence of social norms from costly collective normative efforts exerted by rational individuals, who undertake them because norms legitimize institutional functionality and thereby shape their own payoffs. Institutions are thus not merely functional or coercive mechanisms, but embodiments of shared normative commitments (Acemoglu and Jackson (2017), Bénabou and Tirole (2025)), as long emphasized in the cultural anthropological literature (Geertz (1973), Sahlins (1976), Abeles (1988)). In other words, the standard dichotomy in economics between culture and institutions (see Acemoglu and Robinson (2025) for a recent overview) is conceptually flawed: institutions are not merely shaped by norms; they are normatively constituted through them. This paper formalizes the stabilization of norms as the outcome of institutional and peer normative embodiment. Specifically, it highlights that in the absence of strong institutional embodiment, pluralism – persistent localized normative clusters – can emerge, provided a robust public sphere exists, consistent with Habermas (1989)'s notion of communicative rationality. This perspective may help explain the findings of Ostrom (1990), who documents how communities can self-organize to manage small-scale common resources through informal institutions, precisely because they *autonomously* create and continuously reshape shared social norms within their public sphere. It may also shed light on Bondar and Fuchs-Schündeln (2023)'s finding that East Germans continue to exhibit stronger preferences for redistribution than West Germans, even 27 years after reunification. In contrast, legitimate centralized institutions emerge when a society succeeds in scaling up its shared normative order to the level of the entire polity. This paper shows that such an equilibrium is more likely to be achieved when the instituted social norm is strongly embodied by institutions. Furthermore, in line with Acemoglu and Robinson (2025), it emphasizes that normative orders are not inherently persistent but evolve in response to socio-historical conditions – and can do so rapidly when the public sphere is robust. This paper goes further by identifying autonomy, grounded in creative imagination, as the driving force behind such cultural transformation. This resonates with Mokyr (2016), who argues that the Enlightenment created the cultural foundations for economic transformation in Western Europe by fostering intellectual openness: "For an economy to create the technical advances that enabled it to make the huge leap of modern growth, it needed a culture of innovation, one in which new and sometimes radical ideas were respected and encouraged, heterodoxy and contestability were valued, and novelty tested, compared, and diffused if found to be superior by some criteria to what was there before." (Mokyr (2012), p. 39). Importantly, creative imagination must be distinguished from strategically constructed narratives (Shiller (2017), Bensel (2022)). The *radical imaginary*, as conceptualized by Castoriadis (1987), refers to the human capacity to institute new norms *ex nihilo*. As a corollary, cultural persistence in heteronomous societies reflects not genuine stability but normative rigidity.

This perspective offers a fundamental new insight into the role of states and bureaucracy by identifying them as norm-bearing institutions, that is, institutions that embody the social norm. While most of the economics literature has focused on their functional dimension (see Besley et al. (2022) for a review), models of cultural transmission have highlighted the formative role of specific state agents – teachers – as privileged cultural models (Bisin and Verdier (2001)). Empirical studies underscore the centrality of teachers in shaping national consciousness during the consolidation of modern European nation-states, fostering both language homogenization (Tilly (1975), Hobsbawm (1990), Blanc

and Kubo (2023)) and the diffusion of societal values through national narratives (Weber (1976), Anderson (1983), Clots-Figueras and Masella (2013)). In contrast, this paper conceptualizes, and provides empirical evidence, that states, like all institutions, are not merely functional entities but also symbolic constructs, whose legitimacy is normatively grounded. This view echoes Rousseau's *Social Contract*, in which the state embodies the general will, and aligns with Ekeh (1975)'s *Two Publics*, which, in the African context, distinguishes between morally grounded indigenous institutions and an amoral civic public associated with colonial and postcolonial state structures. This is also consistent with Weber (1922)'s model of legal-rational authority, in which legitimacy rests on citizens' belief in the normative validity of institutional rules, thereby conferring status and moral authority upon bureaucrats. Weber's emphasis on bureaucratic professionalism further implies that, in exchange for security of tenure, bureaucrats should serve the mission of their institution rather than political actors or personal interests. This paper provides evidence that they are more likely than the general population to endorse the established norm and may choose to adopt it as part of their institutional role. It further shows that *all* state agents contribute to shaping the deliberative normative environment of citizens through their embodiment of the state's symbolic dimension. Bureaucracy, then, is neither neutral, alienated, nor obedient (Arendt (1963), Bourdieu (1994), Heldring (2020), Heldring (2023)); rather, its agency acts as an accelerator of convergence toward the instituted norm. This perspective also has important implications for bureaucratic effectiveness: efficiency gains may arise when state employees are motivated by the symbolic mission of their institutional role (Besley and Ghatak (2005), Cassar and Meier (2018), Athias and Wicht (2025)). These gains may be further enhanced by combining material incentives with moral messaging anchored in the instituted social norm, thereby leveraging complementarities between extrinsic and intrinsic motivation (Kranton (2019)). Finally, this paper sheds new light on democratic participation and state capacity. Since the translation of the established norm into institutions' functionality is a condition for autonomy – that is, for collective normative effort – weak translation may undermine democratic participation by eroding the symbolic foundation through which individuals perceive themselves as autonomous agents within a shared normative project, thereby weakening the public sphere (Putnam et al. (1993)). This erosion ultimately reduces state capacity by undermining voluntary compliance (Acemoglu and Jackson (2017), Levi (1997), Tyler (2006)).

This work further contributes to the literature on social preferences. A central element of autonomy as conceptualized in this paper is the notion of layered agency within individuals – conceived both as private agents and as members of a political and social collective. This resonates with the philosophical view that the individual is not ontologically prior to society but rather a product of social-historical institutions (e.g. Castoriadis (1987), Sandel (1998)). Social preferences across these layers may be aligned – or diverge – through deliberative, and even strategic, reasoning (Athias and Ventelou (2025)). Thus, this paper points to a potential confusion in the existing literature, which often conflates private and public-layer social preferences, leading to misinterpretation of the underlying motivations. It also offers a coherent explanation for within-individual variation in social preferences and behavior (Bénabou and Tirole (2011)), showing how the public layer dynamically influences the private layer through self-deliberation shaped by institutional normative embodiment and the public sphere. Overall, incorporating autonomy into the analysis of social preference yields a richer conception of the individual as an agent shaped by, yet capable of reflexively engaging with, social norms. In other words, autonomy en-

tails not only rational deliberation over outcomes but also rational engagement with the normative order itself – that is, the capacity to question, evaluate, and even revise the very norms that govern social life.

My contribution also extends to the literature on universalism by challenging the view that universalism is a fixed personal preference (Cappelen et al. (2025), Enke et al. (2022), Tabellini (2008a)). This paper shows instead that universalism can emerge as a self-instituted social norm and that individuals may *come to endorse* universalist values in the private sphere through deliberation with their environment over a collectively constructed norm. Moreover, in line with Enke et al. (2023), this paper conceptualizes universalism not as an abstract or moral ideal, but as a normative commitment anchored in state functionality – and shows that it is not defined by the size of government, but by a bundle of policy preferences reflecting collective aims. While prior work identifies universal morality as a distinct feature of Western societies, rooted in the Western Church and the Enlightenment (Henrich (2020), Taylor (2004)), this paper emphasizes that its persistence depends on the strength of institutional and peer normative embodiment. Thus, this paper demonstrates that universalism remains fragile and may be reversed under formalized conditions – namely, autonomy combined with a radical exclusionary social imaginary, weakened institutional legitimacy, and a robust public sphere. These conditions closely parallel those associated with the rise of Nazism (Becker and Voth (2025), Mosse (1964), Kohn (1950), Satyanath et al. (2017)). Since institutions embody normative order, inclusive institutions – widely recognized as key drivers of economic prosperity (Acemoglu et al. (2001), Acemoglu et al. (2011), Acemoglu and Robinson (2012)) – can be interpreted as institutional embodiments of universalism as a social norm.

The remainder of this paper is structured as follows. I introduce a simple framework formalizing autonomy in a utility function in Section 2. Section 3 contains details on the data. Section 4 discusses the empirical strategies to test whether universalism is a self-instituted social norm and reports results as well as evidence of mechanism. In Section 5, I leverage the workplace setting to test the state normative legitimacy. Section 6 concludes.

2   The Model

A key insight from Castoriadis (1987) is the concept of *autonomy*. In autonomous societies, individuals recognize social norms as endogenous – created and continually reshaped through normative effort, rather than passively inherited. Autonomy, then, implies that norms do not arise as mere aggregations of individual preferences, but through a collective normative effort, wherein individuals see themselves not only as preference holders but as co-creators of norms. What motivates rational individuals to engage in such costly normative effort? To explore this, I present a simple model of autonomy.

2.1   Model Setup: Autonomy

The instituting social imaginary. We consider an autonomous society represented by a continuum of individuals indexed by $i \in [0,1]$, with total mass normalized to one. Each individual intentionally chooses a normative effort level $e(i,t) \geq 0$, reflecting the deliberate energy, reflection, and collective action they devote to the ongoing production and reshaping of the social norm $Q(t)$ – for instance, through democratic participation, civic activism, or other forms of engaged citizenship. The aggregate social norm $Q(t)$ is endogenously determined by the total normative effort voluntarily exerted in society.

Formally:

$$Q(t) = \phi \left(\int_0^1 e(i,t)\,di \right), \quad \phi' > 0, \quad \phi'' < 0. \tag{1}$$

The autonomy-based nature of the social norm $Q(t)$ lies in the fact that it does not arise mechanically from aggregating individual preferences, but rather emerges from the distributed, voluntary efforts of individuals to construct and sustain a shared normative content that grounds and legitimizes institutions. The model assumes a continuum of individuals, so each agent's contribution $e(i,t)$ has measure zero and no direct marginal impact on the social norm. This implies that agents perceive $Q(t)$ as exogenously given, even though it is endogenously constituted through decentralized participation. This structure reflects Castoriadis (1987)'s view that no individual can claim privileged agency over symbolic meaning, as all meaning arises from the instituting imagination of the collective, the *social imaginary*[7].

The function ϕ, which maps aggregate effort to the social norm $Q(t)$, captures the non-linear returns to normative engagement. Its increasing and concave form reflects that broader participation strengthens the meaningfulness and coherence of the norm, while marginal returns diminish as the effort base expands. Importantly, ϕ also characterizes institutional responsiveness: how effectively institutions internalize and translate collective normative effort into functionality. This reflects Castoriadis (1987)'s view that institutions possess both a symbolic and a functional dimension, and that these dimensions are interdependent: institutional functionality is sustained only through symbolic legitimacy, and symbolic legitimacy must be continually enacted through functional responsiveness.

Overall, this formalization captures Castoriadis (1987)'s distinction between the *instituting* and the *instituted*: the social norm $Q(t)$ is not a pre-existing constraint but the outcome of the instituting social imaginary. Institutions, as embodiments of $Q(t)$, are thus meaningful only insofar as they result from and remain responsive to collective normative participation.

Individual radical imaginary. Each individual $i \in [0,1]$ derives utility at time t from the instituted social norm $Q(t)$, but incurs a personal cost from exerting normative effort $e(i,t)$. Additionally, individuals may experience disutility when the collective norm $Q(t)$ diverges from their normative stance $s(i,t)$. Formally, the utility function of individual i at time t is:

$$U(i,t) = Q(t) - \lambda(i)\big(Q(t) - s(i,t)\big)^2 - c\big(e(i,t)\big), \tag{2}$$

where

- $\lambda(i) \geq 0$ measures the individual's sensitivity to normative misalignment;

- $s(i,t)$ represents the individual's *normative stance* at time t;

- $c(\cdot)$ is a continuously increasing and convex cost function, with $c'(e(i,t)) > 0$ and $c''(e(i,t)) > 0$.

---

[7]This view converges with Sahlins (2022)'s emphasis on the centrality of imagination in creating and sustaining social orders, arguing that societies are not merely pragmatic entities but also works of collective imagination.

Even though each individual has no marginal influence on the social norm $Q(t)$, due to the continuum structure of the population, individuals nonetheless have incentives to exert effort. These incentives arise from the universal benefit of a strong normative order, as captured by the positive term $Q(t)$ in their utility. The social norm legitimizes institutional functionality, thereby fostering voluntary compliance and, in turn, enhancing institutional effectiveness (Acemoglu and Jackson (2017), Levi (1997), Tyler (2006)).

However, autonomy implies that individuals may not be aligned with the instituted norm: their effective normative stance $s(i,t)$ may diverge from $Q(t)$, generating disutility from $Q(t)$ due to alienation costs. This divergence reflects not only a misalignment but the deeper individual capacity for critical distance – the ability to question, reimagine, and potentially reshape social norms – rooted in the radical imaginary (Castoriadis (1987)). This links directly to Castoriadis' core claim: the instituted order is never fixed because the instituting social imaginary can always reopen it. The second term in the utility function captures disutility from such *autonomy-based dissensus*: a deliberative normative gap between an individual's normative stance $s(i,t)$ and the instituted social norm $Q(t)$.

The third term represents disutility from effort.

Layered agency. Autonomy implies that each individual $i \in [0,1]$ possesses two distinct but interrelated normative orientations at time t: a private normative preference $q(i,t)$, representing their personal normative convictions in the private sphere, and a public institutional support indicator $a(i,t) \in [0,1]$, reflecting the degree to which they endorse the institutionalization of their private preference at the societal level. Thus, $a(i,t)$ does not directly measure support for the social norm $Q(t)$, except in the case when $q(i,t) = Q(t)$, where public and private preferences coincide. The product

$$s(i,t) := a(i,t)q(i,t)$$

captures the individual's effective normative stance toward the social norm $Q(t)$.

This formalization captures the idea of *layered agency*, whereby individuals distinguish between their private convictions and their public normative commitments, while recognizing their interdependence. This resonates with Arendt (1958)'s distinction between the private and public realms: the private sphere, associated with necessity, intimacy, and individual life; and the public sphere, where individuals appear before others to deliberate, act, and be held accountable. For Arendt, true political agency emerges in the public realm, where one's words and actions contribute to the shared world. In this sense, the model's distinction between $q(i,t)$ and $a(i,t)$ captures the tension between personal normative reflection and the act of publicly endorsing or institutionalizing those values.

Thus, this layered structure allows for within-individual normative divergence: private preferences may not always translate into the public sphere, reflecting normative restraint, strategic considerations, or broader deliberative tensions inherent to autonomous agency. In what follows, I distinguish several forms of normative divergence that stem from this layered structure:

(ii) **Private endorsement, public rejection** $(q(i,t) = Q(t), a(i,t) = 0)$: The individual privately endorses the prevailing norm but deliberately withholds support for its institutionalization. Two conceptually distinct motivations may underlie this stance:

– *Normative restraint*: The individual refrains from institutionalizing their private preference out of normative concern – for instance, believing that its universal imposition would harm pluralism or social cohesion.

     – *Strategic divergence*: The individual withholds support for instrumental reasons, anticipating that institutionalizing its private preference would reduce their personal payoff – for example, by increasing redistribution or limiting market opportunities.

(iii) **Private rejection, public endorsement** $(q(i,t) \neq Q(t), a(i,t) > 0)$: The individual supports the institutionalization of a norm they do not personally endorse – motivated perhaps by solidaristic commitments, long-run institutional goals, or a belief in collective identity that transcends individual preference.

These *autonomy-based divergences* between an individual's normative orientation in the private and public realms, captured through a concept of *layered agency*, do not imply irrationality or inconsistency. Rather, they reflect a more expansive form of rational agency, in which individuals critically assess not only their personal preferences in their private sphere but also what norms ought to govern collective life. Autonomous agents are capable of such second-order reflection: they may endorse a norm for society that differs from their own preference, either out of normative commitment or strategic consideration, knowing that the instituted norm will shape future payoffs. The resulting configurations of individual normative positions vis-à-vis the instituted norm $Q(t)$ are summarized in Table 1.

Dynamics of normative endorsement. Autonomy also implies that individuals continuously deliberate over their normative stance $s(i,t)$, consistent with Castoriadis (1987)'s conception of the social imaginary, in which social meaning is not fixed or determined, but rather created and continually redefined by society itself. This deliberation does not operate in a vacuum but instead is shaped by their normative environment.

Formally, the evolution of $s(i,t)$ can be expressed as:

$$\frac{\partial s(i,t)}{\partial t} = \underbrace{\alpha\big(Q(t) - s(i,t)\big)}_{\text{institutional embodiment}} + \underbrace{\beta\big(\bar{s}_{-i}(t) - s(i,t)\big)}_{\text{peer embodiment}}, \tag{3}$$

where $\alpha, \beta > 0$ are parameters measuring responsiveness to institutional and peer normative environment, respectively.

The first channel, *institutional embodiment*, reflects a corollary of normative institutional legitimacy: by embodying the social norm $Q(t)$, institutions shape individuals' deliberative environment. For example, when individuals encounter state agents (e.g., social workers, bureaucrats), the experience is symbolically charged and may provoke a reflexive response whereby individuals reconsider and potentially shift their position regarding the social norm $s(i,t)$ – not through coercion but through normative confrontation by a lived institutional representation of the social norm. The parameter α captures the weight of this normative feedback from institutions. A higher α implies greater responsiveness to institutional embodiment, especially for individuals frequently exposed to state agents or living in environments where the state plays a prominent symbolic role via rituals, bureaucratic routines, or the everyday symbolic fabric.

The second mechanism, *peer embodiment*, captures influence from the public sphere via deliberation and discourse with peers. Here, $\bar{s}_{-i}(t)$ denotes the average normative stance of peers in individual i's social environment, and β reflects the strength of public sphere influence.

Table 1: Typology of Individual Normative Positions Regarding the Social Norm $Q(t)$

Case	Private Preference $q(i,t)$	Support for Institutionalization $a(i,t)$	Interpretation
(i)	$q(i,t) = Q(t)$	$a(i,t) \approx 1$	**Normative Harmony**: Individual i endorses the norm in the private and public spheres, showing full alignment between individual and society.
(ii)	$q(i,t) = Q(t)$	$a(i,t) \approx 0$	**Normative Restraint or Strategic Divergence**: Individual i endorses the norm privately but withholds or partially withholds support in the public sphere – either ethically (restraint) or for strategic reasons.
(iii)	$q(i,t) \neq Q(t)$	$a(i,t) \approx 1$	**Normative Alienation**: Individual i rejects $Q(t)$ in the private and public spheres. Experiences disutility from living under a misaligned social order.
(iv)	$q(i,t) \neq Q(t)$	$a(i,t) \approx 0$	**Civic Commitment**: Individual i does not promote the institutionalization of their private preference $q(i,t)$, instead endorsing $Q(t)$ in the public sphere out of civic or solidaristic commitment.

Thus, rather than treating preferences for the social norm as fixed, this structure models preference formation as a dynamic process of normative alignment: individuals continuously adjust their stance $s(i, t)$ in light of normative embodiment from institutions and deliberative engagement with peers. The relative magnitudes of α and β govern the balance between these sources of normative influence. These deliberations constitute the primary channel of normative adaptation. Private preferences $q(i, t)$ may evolve consequentially, as individuals reflect on their societal commitments and update their private normative convictions accordingly.

2.2   Analysis

Proposition 1 (Effort). *Let the utility of agent $i \in [0, 1]$ be given by*

$$U(i, t) = \phi \left(\int_0^1 e(i, t)\, di \right) - \lambda(i) \left(\phi \left(\int_0^1 e(i, t)\, di \right) - s(i, t) \right)^2 - c(e(i, t)),$$

where ϕ is strictly increasing and concave ($\phi' > 0, \phi'' < 0$), c is strictly convex ($c'' > 0$), $a(i), \lambda(i) \geq 0$, and $s(i, t)$ is the preference for the social norm.

Then the optimal effort $e^*(i, t)$ for agent i at time t satisfies:

1. *Effort increases with better alignment between the social norm $Q = \phi\left(\int e(i, t)di\right)$ and the individual's effective normative stance $s(i, t) = a(i, t)q(i, t)$.*

2. *Effort decreases with larger sensitivity to misalignment, measured by $\lambda(i)$.*

3. *Effort exhibits strategic substitutability: as aggregate effort from others increases, the marginal benefit of own effort decreases.*

4. *Effort increases when institutions are more responsive (less concave ϕ), enhancing marginal benefit of effort.*

Proof Sketch.   Define total effort excluding i:

$$E_i := \int_0^1 e(j)dj - e(i).$$

Then the derivative of U_i with respect to $e(i)$ is

$$\frac{dU_i}{de(i)} = \phi'\left(E_i + e(i)\right)\left[1 - 2\lambda(i)\left(\phi(E_i + e(i)) - a(i)q(i)\right)\right] - c'(e(i)).$$

Setting this equal to zero yields the first order condition:

$$\phi'\left(E_i + e(i)\right)\left[1 - 2\lambda(i)\left(\phi(E_i + e(i)) - a(i)q(i)\right)\right] = c'(e(i)).$$

(1) When the social norm $Q = \phi(E_i + e(i))$ closely matches the individual's effective normative stance $s(i) = a(i)q(i)$, the misalignment penalty diminishes, increasing marginal benefit and effort.

(2) A larger sensitivity parameter $\lambda(i)$ increases the penalty weight, lowering marginal benefit and reducing e_i^*.

(3) Because ϕ is concave, increasing aggregate effort E_i lowers ϕ', reducing marginal benefit of own effort and showing strategic substitutability.

(4) More concave ϕ (stronger diminishing returns) reduces marginal benefit, decreasing effort incentives.

$\square$

Discussion. This proposition highlights that while the social norm $Q(t)$ benefits all, individuals' willingness to contribute effort $e(i,t)$ to shape the social norm $Q(t)$ depends critically on their effective normative stance $s(i,t)$. Specifically, when an individual's effective normative stance $s(i,t)$ closely aligns with the currently instituted norm $Q(t)$, they perceive greater benefit in contributing effort because their normative goal and the social norm coincide. Conversely, when $s(i,t)$ diverges from Q, the individual experiences autonomy-based dissensus, reducing their incentive to invest effort in reshaping the norm. Autonomy thereby introduces meaningful heterogeneity in participation. Thus, while $q(i,t)$ does not directly affect $Q(t)$, its interaction with $a(i,t)$ influences individual effort, which shapes $Q(t)$. Furthermore, a higher sensitivity parameter $\lambda(i)$ amplifies this disutility, which can further reduce effort if there is misalignment.

Due to diminishing returns on collective effort captured by the concave function ϕ, individual effort exhibits strategic substitutability: as the aggregate effort $\int_0^1 e(j)\,dj$ grows, the marginal benefit of further effort decreases, tempering incentives to contribute more.

Moreover, the shape of the institutional response function ϕ plays a critical role in effort incentives. Across institutional environments, societies may differ in the responsiveness of institutions to collective normative effort – formally reflected in the curvature of ϕ. A more concave ϕ (i.e., stronger diminishing returns) implies that each unit of effort has less influence on institutional functionality, weakening the perceived efficacy of participation. Consequently, individuals reduce their effort. Conversely, when ϕ is less concave (closer to linear), institutions respond more strongly to individual normative action, increasing the marginal benefit of effort and thereby strengthening incentives for engagement.

This comparative static reveals how institutional design – specifically, the translation of shared normative effort into institutional functionality – can materially affect the sustainability of voluntary norm formation. This insight aligns with Castoriadis (1987), who emphasizes that institutions must be perceived as "instituent" – that is, open to transformation through collective meaning-making – in order to sustain autonomy.

Overall, this framework demonstrates how institutional functional responsiveness and normative alignment jointly sustain autonomous participation in collective norm formation.

Proposition 2 (Normative Stance Steady State). *Let the effective normative stance $s(i,t)$ evolve according to the differential equation*

$$\frac{\partial s(i,t)}{\partial t} = \alpha\big(Q(t) - s(i,t)\big) + \beta\big(\bar{s}_{-i}(t) - s(i,t)\big),$$

where $Q(t)$ is the instituted social norm, $\bar{s}_{-i}(t) = \int_{j \neq i} s(j,t)\,dj$ is the average peer normative stance in individual i's social environment, and $\alpha, \beta > 0$ represent the strengths of institutional normative embodiment and peer deliberation, respectively.

Then the individual steady-state effective normative stance $s^*(i)$ satisfies:

$$s^*(i) = \frac{\alpha Q + \beta \bar{s}_{-i}}{\alpha + \beta}.$$

Interpretation. At steady state, an individual's effective support for the social norm $s^*(i)$ is a convex combination of the institutionalized norm Q and the average peer normative stance $\bar{s}_{-i}$, weighted by the relative influence of institutional normative embodiment (α) and peer deliberation (β). This equilibrium reflects how preferences for the social norm $s(i)$ are shaped by both institutions and social discourse.

Comparative Statics.

- $\frac{\partial s^*(i)}{\partial Q} = \frac{\alpha}{\alpha+\beta} > 0$: Greater institutional normative embodiment leads to stronger alignment with the instituted norm.

- $\frac{\partial s^*(i)}{\partial \bar{s}_{-i}} = \frac{\beta}{\alpha+\beta} > 0$: Stronger peer deliberation increases conformity to peer normative stances.

- As $\alpha \to 0$, $s^*(i) \to \bar{s}_{-i}$: Peer influence dominates, possibly leading to localized normative clusters.

- As $\beta \to 0$, $s^*(i) \to Q$: Institutional normative embodiment dominates, driving convergence toward the social norm.

Discussion. The effective normative stance $s(i,t)$ converges to a weighted average of institutional signals and peer influences. Normative convergence to Q emerges when institutional influence outweighs peer heterogeneity, but pluralism or local divergence may persist otherwise. This reflects how autonomous societies can sustain the social norm not through full normative consensus, but through sufficient public endorsement of the social norm.

Proposition 3 (Endogenous Norm Collapse and Shift)**.** *Consider the dynamic evolution of individual normative stance:*

$$\frac{\partial s(i,t)}{\partial t} = \alpha\big(Q(t) - s(i,t)\big) + \beta\big(\bar{s}(t) - s(i,t)\big),$$

where $s(i,t) = a(i,t)q(i,t)$, $\bar{s}(t) = \int_0^1 s(j,t)\,dj$, and the instituted norm is endogenously determined by:

$$Q(t) = \phi\left(\int_0^1 e(i,t)\,di\right),$$

with individual effort $e(i,t)$ implicitly defined by the first-order condition:

$$\phi'\left(\int_0^1 e(i,t)\,di\right)\big[1 - 2\lambda(i)\big(Q(t) - s(i,t)\big)\big] = c'\big(e(i,t)\big).$$

In regimes where institutional normative embodiment is weak relative to peer deliberation ($\alpha \ll \beta$), the average normative stance $\bar{s}(t)$ becomes the dominant influence on individual updates. This dynamic structure generates the following feedback mechanisms:

- ***Endogenous Norm Shift****: Small perturbations in peer preferences – such as those introduced by the activation of a radical imaginary, i.e., individuals' autonomous creation of new norms outside existing instituted norms – can shift $\bar{s}(t)$ through deliberative interaction. This may lead to gradual realignment of the entire preference distribution $s(i,t)$ away from the instituted social norm $Q(t)$ in the absence of sufficient institutional stabilization (i.e., low α).*

- **Effort Withdrawal Feedback**: *Since individuals' effort levels $e(i,t)$ depend on the alignment between $s(i,t)$ and $Q(t)$, increasing misalignment reduces contributions to the norm, i.e., $e(i,t)$ decreases, which in turn lowers $Q(t)$, initiating a reinforcing feedback loop.*

- **Normative Collapse and Shift**: *As voluntary support and effort decline, the instituted norm $Q(t)$ erodes endogenously, resulting in symbolic collapse from below and the emergence of a new symbolic order.*

The model highlights how symbolic collapse can arise endogenously from decentralized peer deliberation and weakened institutional anchoring – paving the way for reconfiguration through the radical imaginary.

Illustration. The rise of Hitler in the 1932 German elections can be interpreted through the lens of a norm collapse driven by peer dynamics, as captured in the above proposition. In the late Weimar Republic, state institutions suffered from legitimacy crises and lacked normative embodiment, while the public sphere remained vibrant (Satyanath et al. (2017)). This corresponds to the model configuration where institutional normative influence is weak relative to peer influence: $\alpha \ll \beta$. This allowed β-driven peer dynamics to fill the void, resulting in a cascading shift of individual preferences and the eventual collapse of the existing normative order.

Crucially, this transformation did not require an exogenous shock; it emerged endogenously from the fragility of the symbolic order when institutional normative embodiment was weak and the radical imaginary in the public sphere was important. The radical imaginary – embodied in fascist mythologies of rebirth and unity[8] – emerged and spread through decentralized peer interactions, ultimately eroding the instituted norm Q. Hitler's electoral success can thus be seen not as an imposition from above, but as the outcome of a decentralized norm collapse.

3 Data

To empirically evaluate the model, I focus on Switzerland – a country situated within post-Enlightenment Western Europe, where universalism is generally expected to be the prevailing social norm (Henrich (2020), Taylor (2004)). However, Switzerland's federal structure, strong local identities, and history of selective integration could introduce meaningful variation in the extent to which this norm is endorsed. In addition, Switzerland's semi-direct democratic institutions, political neutrality, tradition of pragmatic consensus, and institutional stability make it a particularly informative case for assessing the degree to which autonomy is realized in practice.

The data come from the Swiss Household Panel Surveys (SHPS), a unique longitudinal survey conducted annually since 1999. The SHPS interviews all members of a random sample of approximately 5,000 households, covering over 12,000 individuals residing

---

[8]See Mosse (1964) and Kohn (1950) on how 19th-century Romanticism fostered a cultural climate emphasizing racial purity and rejecting Enlightenment universalism – ideas diffused across the German population prior to 1914. More recent evidence from PewResearchCenter (2011) reveals substantial cross-country variation in cultural imaginaries. Approximately half of Americans (49%) and Germans (47%) agree with the statement, "Our people are not perfect, but our culture is superior to others." In Spain, 44% of respondents share this view, compared to about a third of Britons (32%) and 27% of French respondents.

in Switzerland. It collects rich individual-level information on demographic and socio-economic characteristics, along with a broad range of attitudinal and behavioral topics. I use the SHP individual-level panel data from waves I (1999) to V (2003), the only waves that contain all the necessary variables. I restrict the sample to Swiss individuals whose first nationality is Swiss – since the question used to elicit preferences for universalism as a social norm involves a trade-off between Swiss citizens and foreigners – and who are over 18 years old, that is, beyond their formative years. The data are geolocated at the municipality level, corresponding to respondents' commune of residence. My final sample contains 7'125 individuals (22'618 person observations).

The surveys include a precise measure of individuals' preference for the universalism social norm, based on the following question: "Are you in favour of Switzerland offering foreigners the same opportunities as those offered to Swiss citizens, or in favour of Switzerland offering Swiss citizens better opportunities?". For Swiss respondents, this question presents a clear trade-off between equal treatment of in-group members (Swiss citizens) and socially distant others (foreigners), versus preferential treatment of the in-group. It thereby directly elicits their preference for universalism – as defined by Enke et al. (2022) in terms of private social preference – but here applied to a society-wide norm[9]. The three response options are: "in favour of equality of opportunities", "neither", or "in favour of better opportunities for Swiss citizens". I construct the individual-level variable $UNIV$ as a binary indicator coded as 1 for respondents who endorse the universalism social norm, and 0 for those who prefer preferential treatment for Swiss citizens. Respondents selecting the "neither" option are excluded from this baseline measure. This exclusion is justified both empirically and conceptually: the "neither" category comprises less than 10% of the sample and likely captures heterogeneous responses, including ambivalence, measurement error, or social desirability bias, rather than a coherent normative position. Restricting the analysis to a dichotomous measure also facilitates interpretation in subsequent analyses, particularly when examining deliberative and treatment effect heterogeneity. To assess the sensitivity to this coding choice, I also construct a three-category ordinal variable coded –1, 0, and 1 – where 1 corresponds to respondents endorsing the universalism social norm, 0 to those responding "neither", and –1 to those favoring preferential treatment for Swiss citizens.

Over the full period for which the variable is available (annually from 1999 to 2009, and additionally in 2011, 2014, and 2017), Figure 8 in Appendix A.1 shows that the overall share of individuals endorsing universalism remains relatively stable, though it exhibits a decline following the 2007-2008 financial crisis. In the main estimation sample (1999–2003), 34.3% of individuals do not endorse universalism as a social norm. Figure 1 reveals substantial spatial heterogeneity in support for universalism across cantons, ranging from 51% to 73%, indicating the existence of localized normative clusters. In my model, this pattern is consistent with a high value of β, reflecting a strong role for the public sphere in shaping self-deliberation. Furthermore, Figure 2 plots the within-canton-year standard deviation of individual responses, with a LOESS-smoothed curve tracing the overall temporal trend. The stability of normative dispersion over time suggests that the strength of self-deliberation mechanisms – captured by the model parameters α and

---

[9]Decisions involving trade-offs between in-group members and socially distant others are central to the concept of universalism. Enke et al. (2022) develop a measure of universalism using survey-based money allocation games, in which participants divide a hypothetical sum of $100 between an in-group member and a randomly selected stranger. The defining feature of universalists is not greater morality or generosity per se, but a preference for equal treatment – allocating the budget uniformly across individuals regardless of group affiliation.

β in Equation 3 – remained approximately constant throughout the observation period. Finally, Figure 9 in Appendix A.1 reports ordinary least squares estimates of universalism endorsement on individual-level covariates. Endorsement is positively associated with being college-educated, left-leaning, French-speaking, urban, married, or high-income, and negatively associated with religiosity, being female, or having more children. Importantly, 33% of individuals (26.2% over 1999–2003) change their normative stance at least once. Taken together, the joint stability of both the mean and the dispersion of universalism endorsement – combined with substantial individual-level normative mobility – suggests that the norm is in a state of *dynamic equilibrium*: while aggregate endorsement remains stable, individuals continue to deliberate and revise their normative positions. In the framework of Equation 3, this pattern implies that the self-deliberation parameters α and β are sufficiently balanced to generate micro-level dynamism without macro-level erosion.

Figure 1: Universalism Endorsement Averaged at the State (Cantonal) Level
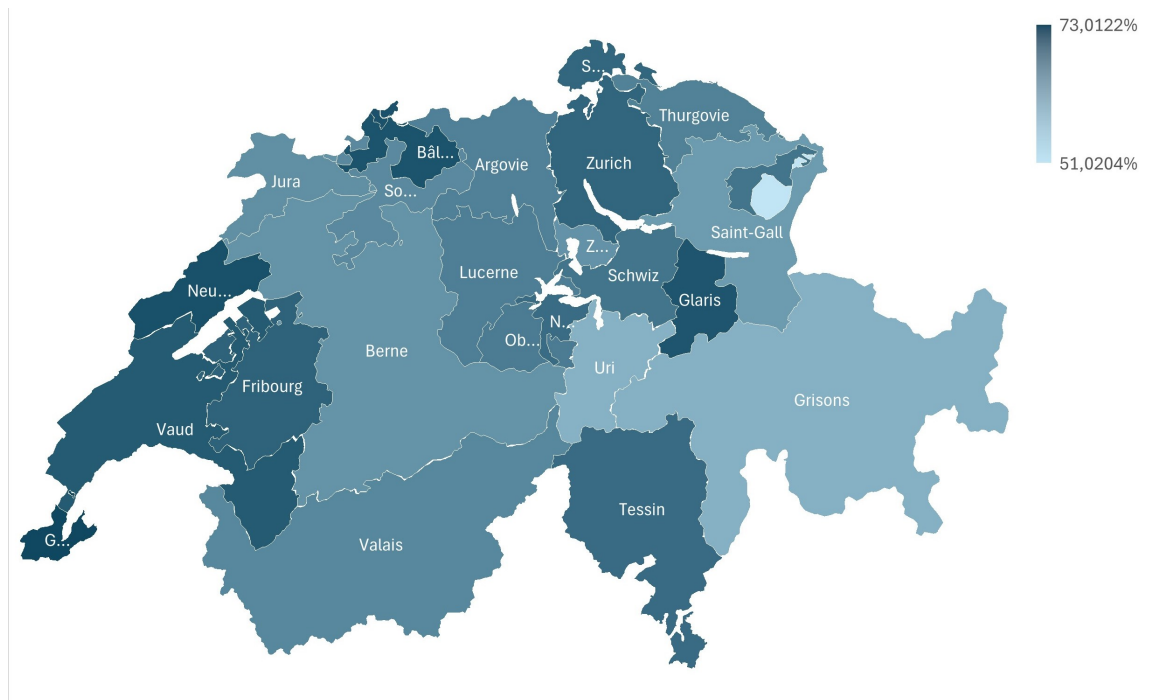
Note: The map displays the average share of Swiss individuals aged 18 and over in each canton who endorse the universalism social norm (binary variable). This corresponds to the cantonal average of the variable $UNIV$ over the period 1999–2017.

I further define individuals' institutional sector based on the following survey question: "Are you employed by a private company or a state organization?". I construct then the binary variable $State$, which equals 1 if the respondent works for the state and 0 if they are employed in the private sector. Respondents are also prompted with a list of options within state organizations: International organization, Confederation/Swiss Railways/Post office, Canton, Commune. Respondents who indicate working for an international organization are excluded, as these are non-governmental entities and not part of the domestic state apparatus. The SHP data also include information about the occupational choice at the service level. Specifically, I use respondents' choice in the nomenclature of economic activities to distinguish workers in public services, defined as health, education and social care in the literature, from the others.

17

Figure 2: Normative Dispersion Over Time
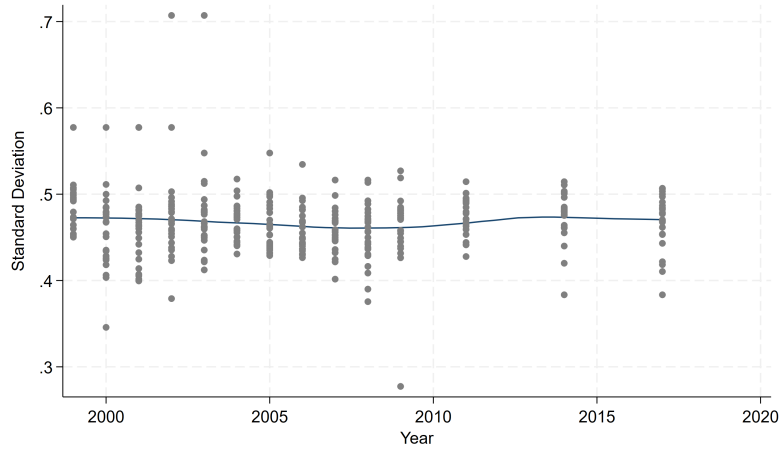
Note: Dots represent the standard deviation of individual responses to the universalism social norm question, computed within each canton and year among Swiss individuals aged 18 and over. The black trend line depicts a LOESS-smoothed curve, capturing the overall temporal trend in normative dispersion.

Table 2 shows a pronounced gap in the endorsement of the universalism social norm between institutional sectors: 74% of state employees support the universalism social norm, compared to 66% in the private sector. This difference reflects a general institutional effect, indicating that the association between public sector employment and universalism endorsement is not confined to traditionally mission-oriented fields. When disaggregating by industry, the gap between state and private sector workers persists within both health, education, and social care (HES) industries and non-HES industries, with similar magnitudes. Notably, the standard deviation of universalism endorsement is consistently lower among state employees, suggesting a more homogeneous normative stance within the public sector. These patterns provide preliminary evidence of the symbolic or normative legitimacy of the state.

However, support for the universalism norm is significantly higher among state employees at the cantonal level (78%) than at the federal level (67%), with a lower standard deviation as well (0.414 versus 0.471). This within-state heterogeneity suggests meaningful variation in how state normative legitimacy is embodied across levels of government in Switzerland – a pattern consistent with Ekeh (1975)'s theory of *Two Publics*, whereby local state institutions are more closely aligned with citizens' moral expectations than central ones. In the Swiss context, this may reflect the limited authority of the federal government relative to the cantons, which hold substantial discretion across most policy domains. Compared to many other countries, Switzerland's federal state institutions are weaker than their cantonal counterparts, potentially shaping the locus of normative legitimacy.

Figure 3 reveals that time trends in universalism endorsement differ markedly between public and private sector employees. While both groups exhibit increases around the year 2000, only the trend among public sector workers continues to rise before stabilizing, whereas the private sector trend flattens and slightly declines. This divergence supports the identification strategy pursued in the next section – specifically, the empirical sep-

18

Table 2: Universalism Endorsement in the Full Sample and by Employment Institutional Sector (1999–2003)

|	Full sample			Public sector		Private sector		
	Mean	N	% ever change UNIV	(1) Mean	N	(2) Mean	N	(3) Difference
All individuals	0,657 (0.475)	22,618	26.2%	0.742 (0.438)	4,250	0.657 (0.475)	7,473	0.085*** (0.009)
Within HES industries				0.800 (0.400)	2,112	0.769 (0.421)	924	0.030* (0.016)
Within non-HES industries				0.666 (0.472)	1,569	0.641 (0.480)	5,265	0.026* (0.014)

Note: % ever change UNIV refers to the share of individuals who change their endorsement of the universalism social norm at least once between 1999 and 2003. Sectoral breakdowns only include employed individuals; full sample statistics include all Swiss individuals aged 18+. HES refers to health, education and social care. Non-HES refers to all other industries. Column (3) reports the difference in mean endorsement of the universalism social norm between public and private sectors. Standard deviations are in parentheses below means in Columns (1) and (2); robust standard errors are in parentheses in Column (3). *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

aration of $UNIV_{ct}^{State}$ and $UNIV_{ct}^{Peers}$ in equation (5) – as these two variables capture distinct normative environments rather than a common temporal trajectory. Moreover, the fact that non-workers' trajectory closely follows that of state employees suggests that institutional normative embodiment (α) plays a central role in shaping their normative orientation. This pattern is consistent with the model's prediction that individuals who are more frequently exposed to public agents – outside of formal employment contexts – are more responsive to the symbolic dimension of the state. Unlike employed individuals who are embedded in specific occupational environments, non-workers often interact with the state as recipients of public services (e.g., unemployment benefits, healthcare, or administrative support), placing them in repeated contact with public agents and the values they embody in their function. As such, the alignment in trends between non-workers and state employees supports the empirical relevance of α in the process of universalism as a self-instituted norm. Conversely, the divergence between public and private sector trajectories underscores the importance of peer deliberation (β) within the working population, where embeddedness in different normative milieus likely drives heterogeneity in the evolution of universalism endorsement.

4 Universalism: a Self-Instituted Norm

In this section, I empirically test the core assumptions and propositions of the theoretical model. I begin by examining whether individuals' endorsement of the universalism norm is shaped by their deliberative institutional and peer normative environments. I then investigate heterogeneity in these effects to uncover the layered agency that characterizes autonomy. Finally, I test whether alignment with the norm increases normative effort, as reflected in patterns of democratic participation and collective engagement.

Figure 3: Yearly share of individuals who endorse the universalism social norm, by group
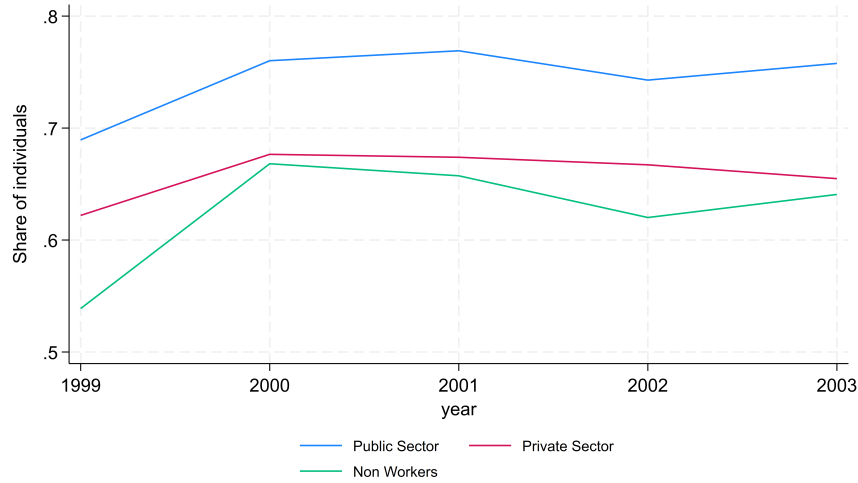
Note: The figure shows the share of Swiss individuals aged 18 and over who endorse the universalism social norm over time, presented as the yearly average of the variable $UNIV$, separately for non workers, state employees, and private sector workers.

4.1 State and Peer Norms Shape the Dynamics of Individual Universalism

To investigate the dynamics of universalism as a self-instituted norm, I exploit longitudinal variation in the normative environment, proxied by the share of state employees and peers endorsing the universalism social norm. Specifically, I examine whether an increase in the cantonal average share of universalist public bureaucrats and peers induces an increase in the likelihood that non-workers in the same canton endorse the universalism social norm. I estimate the following individual-level fixed effects specification:

$$UNIV_{ict} = \alpha \times UNIV_{ct}^{State} + \beta \times UNIV_{ct}^{Peers} + \psi \times Left_{ct} + \delta_t + \lambda_c + \gamma_i + \epsilon_{ict}. \quad (4)$$

The unit of observation is an individual (i) aged 18 or older who is not in employment, $\times$ locality (canton) (c) $\times$ year (t) cell. The dependent variable $UNIV_{ict}$ is an indicator equal to 1 if the individual endorses the universalism social norm, and 0 otherwise. The key explanatory variables are $UNIV_{ct}^{State}$ and $UNIV_{ct}^{Peers}$, which correspond to the yearly average level of universalism endorsement among state employees and private sector workers in the same canton and year, respectively. These variables are computed using individuals' public or private sector of employment based on their commune of residence, not workplace. While this may raise concerns about measurement validity, the potential bias is likely to be small. Commuting in Switzerland is predominantly local: according to the Federal Statistical Office, the average commuting distance in 2000 was 12,9 km; 36.3% of commuters worked within within their commune of residence, 49% in another commune within the same canton, and only 14.7% outside their canton. This is consistent with Eugster and Parchet (2019), who use individual data from the 2000 Federal Population Census to estimate commuting distances for all employed individuals in the three bilingual cantons. They find that more than 80% of individuals reside within

20 kilometres of their workplace. The coefficient α captures the deliberative effect of state normative embodiment, while β captures the deliberative effect through the public sphere. I include individual fixed effects γ_i, canton fixed effects λ_c, and year fixed effects δ_t. To account for time-varying political ideology of the state, I control for $Left_{ct}$, an indicator equal to 1 if the weighted average ideology of the cantonal executive lies to the left of the national average. This variable is constructed from data in Walter and Emmenegger (2019), using the yearly number of cabinet seats held by parties mapped onto a left–right scale from 1 (far left) to 4 (far right). Standard errors are clustered at the individual level. I estimate a linear probability model with individual fixed effects to account for time-invariant unobserved heterogeneity[10].

The results presented in Table 3 indicate that the coefficient on the share of universalists among state employees is positively associated with the endorsement of the universalism social norm among non-working individuals. In Column (1), I find that a one-unit increase in the share of state employees endorsing universalism in a given canton-year leads to a 15.2 percentage-point increase in the probability that a non-worker in the same canton and year endorses the social norm. This corresponds to a 1.52 percentage-point increase in probability for every 10 percentage-point increase in universalist state employees. Given that the mean of $UNIV$ among non-workers is 62,4%, this represents a modest yet meaningful relative increase of approximately 2.4%. This estimate remains positive and of similar magnitude when controlling for cantonal government ideology in Column (2), suggesting that the normative influence of public employees on non-workers' normative stance operates through their institutional role rather than reflecting prevailing left-wing political environments. Column (3) adds the share of private sector workers (peers) who endorse universalism as the additional source of normative deliberation. The coefficient on state employees remains positive and statistically significant, though slightly attenuated. These results corroborate the model's assumption that by embodying the social norm, institutions shape individuals' deliberative environment. Individuals endorse the universalism social norm not due to coercion but because they are normatively confronted by a lived institutional representation of the social norm. Importantly, the peer normative embodiment coefficient is also positive and statistically significant, with a value of 0.176, suggesting that both state and peer normative embodiment independently shape deliberative norm adoption among non-workers. The fact that the estimated coefficients on institutional and peer normative embodiments are of similar magnitude is notable. Given that non-workers are likely to be in frequent contact with state agents through service provision, the slight predominance of the peer channel suggests that deliberation in the public sphere plays a comparatively stronger role in shaping individual normative positions in the broader population. This finding aligns with the spatial distribution of universalism documented earlier, which reveals localized normative clusters. At the same time, the observed empirical stability of both the mean and the dispersion of universalism

---

[10]As robustness checks, I estimate a random-effects logit model, exploiting the binary nature of the outcome variable. The logit model provides a nonlinear alternative under the assumption that individual effects are uncorrelated with the regressors. Given that the main covariates vary only at the canton-year level, this assumption is plausible. The results are fully consistent in sign and statistical significance (see Table 8 in Appendix A.2). In addition, I estimate models using the three-category ordinal outcome (–1, 0, 1) while maintaining the canton-year normative environment variables computed from the binary classification. This specification preserves conceptual clarity in the measurement of normative environments while allowing for greater heterogeneity in individual responses. The results remain fully consistent with the baseline binary specification, with all key coefficients retaining their statistical significance and magnitude (see Table 9 in Appendix A.2).

endorsement suggests that the estimated value of α for non-workers is likely representative of the broader population. Taken together, the total estimated effect of the normative environment – summing state and peer effects – amounts to 0.309. This implies that a simultaneous 10 percentage-point increase in the share of universalist state employees and private sector peers causally increases the probability that a non-worker endorses universalism by 5%. This magnitude highlights the causal importance of the normative environment in shaping self-deliberation over the social norm, consistent with the model's autonomy-based framework.

Table 3: Effect of State and Peer Norms on Universalism Endorsement among Non-Workers

	(1)	(2)	(3)
	Individual Universalism Endorsement		
Share UNIV among state employees	0.152**	0.147*	0.133*
	(0.0752)	(0.0755)	(0.0754)
Left-wing state		0.0144	0.0106
		(0.0257)	(0.0259)
Share UNIV among peers			0.176**
			(0.0794)
Observations	10,842	10,842	10,841
Number of individuals	4,576	4,576	4,575

Notes: All regressions are estimated using ordinary least squares (OLS) and include individual fixed effects, as well as canton and year fixed effects. The dependent variable is a binary indicator equal to 1 if the non-working individual endorses the universalism social norm, and 0 if they favor preferential treatment for Swiss citizens. *Share UNIV among state employees* and *Share UNIV among peers* refer to canton-year averages of universalism endorsement among state and private sector employees, respectively. *Left-wing state* is a binary indicator equal to 1 if the cantonal executive is left-leaning (see text for coding details). Standard errors are clustered at the individual level. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

One potential concern is that the variable capturing left-wing cantonal government ideology may not fully account for the influence of broader political orientations on non-workers' preferences – particularly if this group, often economically or socially vulnerable, is more sensitive to shifts in public policy or rhetoric. A second, related concern is that my results may reflect a general responsiveness of non-workers to the surrounding institutional and social environment, rather than a specific responsiveness to universalism as a social norm. For instance, changes in the normative environment could shape other correlated preferences – such as views on redistribution, inequality, or political institutions – thereby confounding the interpretation of α and β as norm-specific channels. To assess these concerns, I re-estimate the individual-level fixed effects specification from Equation (5), replacing the dependent variable with a series of alternative societal preferences. These include: benevolence (measured by support for decreasing, maintaining, or increasing welfare spending); inequality aversion (measured by preferences regarding taxation of high incomes); views on defense and the environment (measured by support for a strong Swiss army and for environmental protection over economic growth, respectively); political ideology (measured on a 0–10 left–right scale); and confidence in political institutions, proxied by trust in the federal government and satisfaction with democracy (both measured on 0–10 scales). Results are reported in Table 4. Strikingly, the normative composition of the state and peer environment does not significantly affect any of the alternative societal preferences considered, with the sole exception of trust in the federal government. In that case, a more universalist institutional embodiment leads to a decline in trust. This finding is, however, consistent with the previously documented

heterogeneity within the state itself – specifically, the lower endorsement of universalism among federal bureaucrats relative to their cantonal counterparts. Taken together, the absence of parallel effects on related societal preferences reinforces the interpretation that the estimated coefficients on α and β reflect a norm-specific effect on universalism, rather than a broader shift in societal preferences. Furthermore, the positive and statistically significant coefficient on the left-wing state indicator for redistributive preferences suggests that this control does capture broader ideological and policy influences. This strengthens the claim that the identified state normative embodiment effect is not driven by residual confounding from general political orientation.

Table 4: Effect of State and Peer Norms on other Societal Preferences among Non-Workers

	(1) Welfare	(2) Redistri.	(3) Environ.	(4) Army	(5) Pol. ideology	(6) Trust fed. gov.	(7) Sat. demo.
Share UNIV among state employees	-0.0046	0.0089	-0.009	-0.161	0.0890	-0.582**	-0.0498
	(0.115)	(0.104)	(0.118)	(0.106)	(0.290)	(0.285)	(0.284)
Left-wing state	0.0668	0.0685**	0.0411	0.0400	-0.161	0.0256	-0.0743
	(0.0421)	(0.0343)	(0.0411)	(0.0367)	(0.113)	(0.109)	(0.120)
Share UNIV among peers	-0.116	-0.0271	0.0362	0.0251	-0.0453	0.373	0.384
	(0.118)	(0.0990)	(0.115)	(0.106)	(0.280)	(0.289)	(0.291)
Observations	12,033	12,129	12,235	12,273	10,968	12,360	12,321
Number of individuals	4,765	4,770	4,788	4,791	4,517	4,804	4,801

Notes: All regressions include individual fixed effects and canton and year fixed effects. *Share UNIV among state employees* and *Share UNIV among peers* refer to canton-year averages of universalism endorsement among state and private sector employees, respectively. *Left-wing state* is a binary indicator equal to 1 if the cantonal executive is left-leaning (see text for coding details). Outcome variables are described in the main text and correspond to preferences over welfare spending, redistribution, environmental protection, military defense, political ideology, and trust in democratic institutions. Standard errors are clustered at the individual level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

4.2 Evidence of Mechanism: Autonomy as Layered Agency

The theoretical framework posits that the effects of the state and peer normative environments do not arise from passive value transmission or social pressure, but from active processes of deliberation and norm embodiment. Accordingly, the absence of spillovers to other societal preferences supports the interpretation that the estimated parameters α and β capture norm-specific deliberative dynamics, rather than value transmission[11].

To further probe the deliberative nature of the state and peer normative effects, I examine heterogeneity in their impact to uncover the layered agency that characterizes autonomy – that is, the possibility that individuals' endorsement of the social norm diverges from their social preference in their private sphere. Building on Enke et al. (2023), who show that universalism as a private value is the strongest predictor of left-wing ideology, I assess whether the effects of the normative environment vary across the ideological spectrum. This exercise speaks directly to the model's conception of autonomy as layered agency. If autonomy operates through such layering, the institutional and peer normative environment may shape individuals' deliberative endorsement even when they already

---

[11]If the effects reflected value transmission, one would expect lagged measures of the normative composition of the institutional and peer environments to exhibit similar or even stronger influence. However, when using lagged values, the estimated coefficients become statistically insignificant and even negative, suggesting that state and peer normative environments operate primarily through *contemporaneous deliberative interaction*.

hold universalism as a private value but do not endorse it as a societal ideal. Observing stronger or equally strong effects among these individuals would provide empirical support for autonomy: they engage with the norm as a symbolic collective construct, distinct from their private convictions, and revise their normative stance accordingly.

To explore this, I re-estimate the main fixed-effects model with interaction terms between, in two separate specifications, state and peer universalism and indicators for ideological orientation. Ideology is measured on a 0–10 scale, and I construct four bins: left-wing (0–4), centrist (5), right-wing (6–8), and extreme-right (9–10). Figure 4 plots the estimated coefficients on $UNIV_{ct}^{State}$ and $UNIV_{ct}^{Peers}$ for each group, using centrists as the reference category. The results show that both normative environments – state and peer – have the strongest effects among left-wing individuals. This is consistent with the model's concept of layered agency: even those who already hold universalism as a private value may further align their societal stance when deliberating with a more supportive normative environment, revealing normative divergence across agency layers. Interestingly, the effects are also larger among right-wing individuals relative to centrists. This pattern points to *strategic normative divergence* among centrists – a core type of autonomy-based dissensus in the model (see Case (ii) in Table 1): individuals may privately endorse universalism but withhold societal support when they perceive that its institutional implementation threatens their interests – for instance, via redistribution or prioritization of environmental protection. By contrast, both left-wing and right-wing individuals appear more responsive to deliberative normative environments because their divergence from the social norm is not strategic but normative. For left-wingers, norm endorsement reflects an extension of their normative boundaries (the initial divergence stems from *normative restraint*); for right-wingers, norm endorsement may reflect a *civic commitment* to the social norm, even when it diverges from their private views (see Cases (ii) and (iv) in Table 1). This strategic divergence among centrists may also help explain why no significant difference emerges between centrists and individuals on the extreme right. While extreme-right individuals may reject universalism both in the private and public realms on normative grounds, centrists may privately endorse it but withhold societal support out of self-interest. As a result, the deliberative response to normative exposure appears similarly muted across both groups – albeit for different underlying reasons.

To probe this interpretation, I examine whether the universalism norm endorsement correlates with changes in policy preferences within ideological groups. Figure 5 shows that, consistent with Enke et al. (2023), the left-right divide on redistribution strongly depends on whether individuals endorse universalism. Right-wing individuals who adopt the universalism norm shift toward greater support for redistribution – expressing their civic commitment – whereas norm adoption does not affect redistribution preferences among left-wingers or centrists. This pattern confirms that left-wingers and centrists likely already hold universalism as a private preference. However, I also find that norm endorsement leads to a significant decline in support for a strong army among left-wing individuals. This confirms their layered divergence, rooted in normative constraints. As Cappelen et al. (2025) (p.5) note: "A strong military is in some ways an antidote to universalism because it serves to defend boundaries between 'us' and 'them'." In addition, even more pronounced shifts in policy views are observed among centrists: they not only become less supportive of a strong military but also increasingly prioritize environmental protection over economic growth. This indicates that deliberation over the social norm leads them to revise their policy preferences, even in areas where doing so may go against

their material interests, reflecting a deliberative shift in their willingness to incur personal costs in support of institutional legitimacy.

Overall, these findings provide empirical support for autonomy as the mechanism through which state and peer normative environments influence individual-level dynamics of universalism endorsement. They also validate a central assumption of my theoretical model: that the instituted social norm is not an abstract ideal, but legitimizes institutional functionality – thereby affecting individual utility.

Figure 4: Ideological Heterogeneity in the Effects of State and Peer Normative Environments on Universalism Endorsement
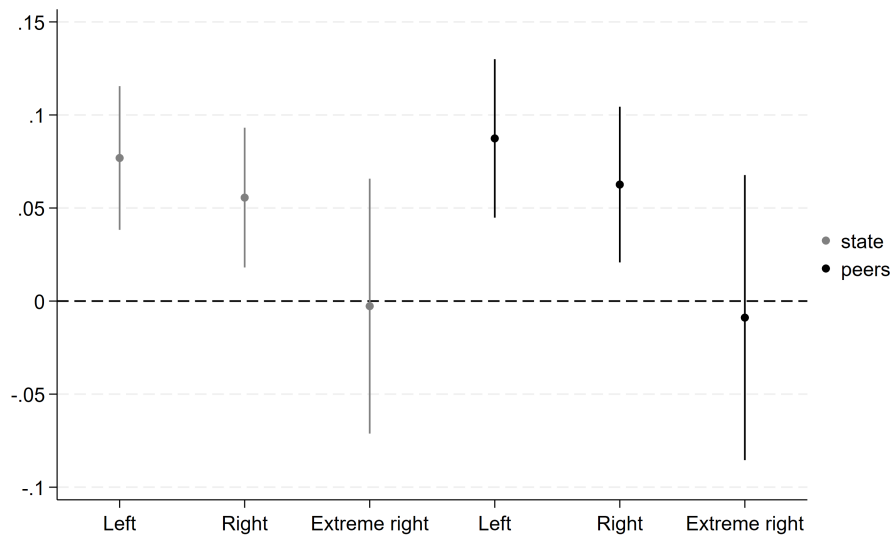
Notes: The figure presents estimates from two separate specifications based on Equation (5), allowing the coefficients on $UNIV_{ct}^{State}$ (grey) and $UNIV_{ct}^{Peers}$ (black) to vary with individuals' ideological orientation. The ideological variable ranges from 0 ("left") to 10 ("right"), and I construct four bins: left-wing (0–4, 32.9% of the non-workers sample), centrist (5, 36.4%), right-wing (6–8, 26.4%), and extreme right-wing (9–10, 4.4%). The centrist category is omitted and serves as the reference group; coefficients for the Left, Right, and Extreme Right ideological groups represent deviations from this baseline. The sample includes non-workers aged 18 and above. Reporting 90% confidence intervals based on standard errors clustered at the individual level.

4.3 Normative Alignment and Democratic or Collective Engagement

To test **Proposition 1** of the model – namely, whether normative effort increases with alignment to the instituted social norm – I examine whether a shift in norm endorsement is associated with changes in behavioral democratic and collective outcomes. The model predicts that individuals who endorse the norm will be more likely to engage in actions that contribute to its continual redefinition, such as democratic participation, civic activism, or other forms of collective engagement. Therefore, I estimate the following individual fixed-effects model:

$$Outcome_{ict} = \theta \times UNIV_{ict} + X_{ct}'\delta + \gamma_i + \epsilon_{ict} \qquad (5)$$

where the unit of observation is an individual (i) age 18 or older in canton (c) and year (t). γ_i denotes individual fixed effects, and X_{ct} is a vector of canton-by-year fixed effects.

Figure 5: Effect of Universalism Endorsement on State-Related Policy Preferences by Ideological Group
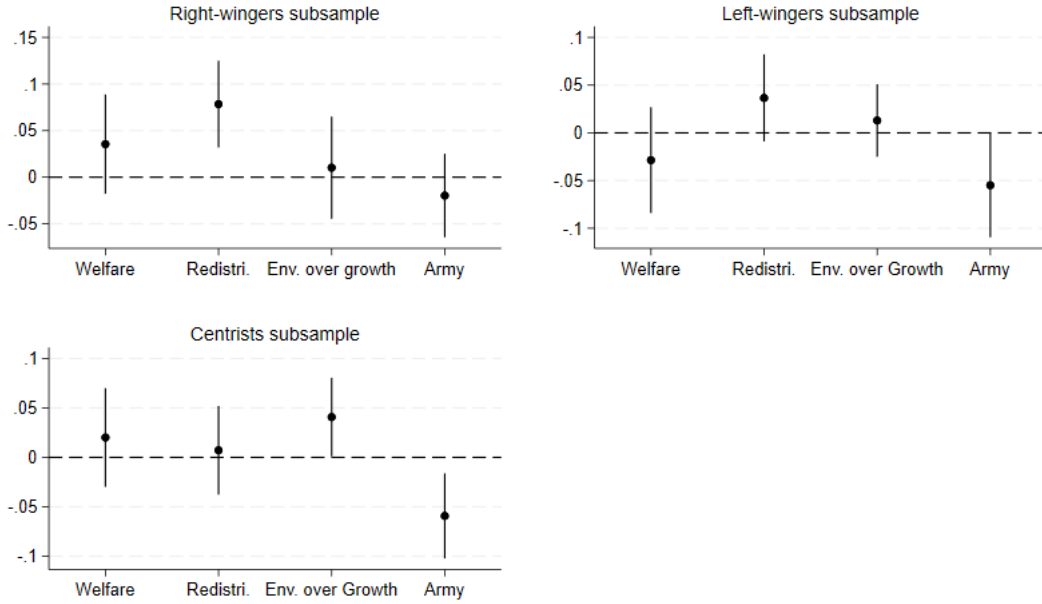
The key independent variable, $UNIV_{ict}$, equals 1 if the individual endorses the universalism social norm in year (t), and 0 otherwise. The dependent variable $Outcome_{ict}$ varies across specifications and captures distinct forms of democratic participation and collective engagement. These include: participation in federal polls; engagement in future demonstrations and strikes; perceived influence over government policy; overall satisfaction with democracy (each measured on a scale from 0 to 10); and binary indicators for membership in an environmental protection or charitable association, as well as participation in clubs or other civic groups.

While the theoretical model implies a causal relationship between norm alignment and normative effort, the empirical analysis identifies patterns of association that are consistent with this mechanism. Given the absence of an exogenous source of variation in alignment, the evidence should be interpreted as descriptive but supportive of the causal structure embedded in the model.

Figure 6 presents the relationship between norm endorsement and democratic or collective engagement. Consistent with the model, individuals who endorse the universalism norm are more likely to participate in federal polls, report greater perceived political influence, and express higher satisfaction with democracy. While not statistically significant, the coefficients on demonstration and strike participation are suggestive: norm endorsement is positively associated with demonstrations but negatively associated with strikes – a pattern consistent with engagement motivated by collective rather than particular interests. This interpretation is further supported by the results for collective engagement variables: all coefficients are positive and statistically significant. Endorsement is associated with greater participation in clubs and associations devoted to collective causes,

such as environmental protection and charitable activities.

These findings support the model's central prediction: when individuals align with the instituted norm, they perceive greater benefit from contributing to its ongoing redefinition, as their normative goal coincides with the social norm. This alignment reduces misalignment costs and increases normative effort. More broadly, the results corroborate the autonomy-based account of norm formation: individuals engage in democratic and collective behavior not only as passive norm-followers, but as co-creators of social norms from which they derive utility.

Figure 6: Universalism Endorsement and Democratic or Collective Engagement
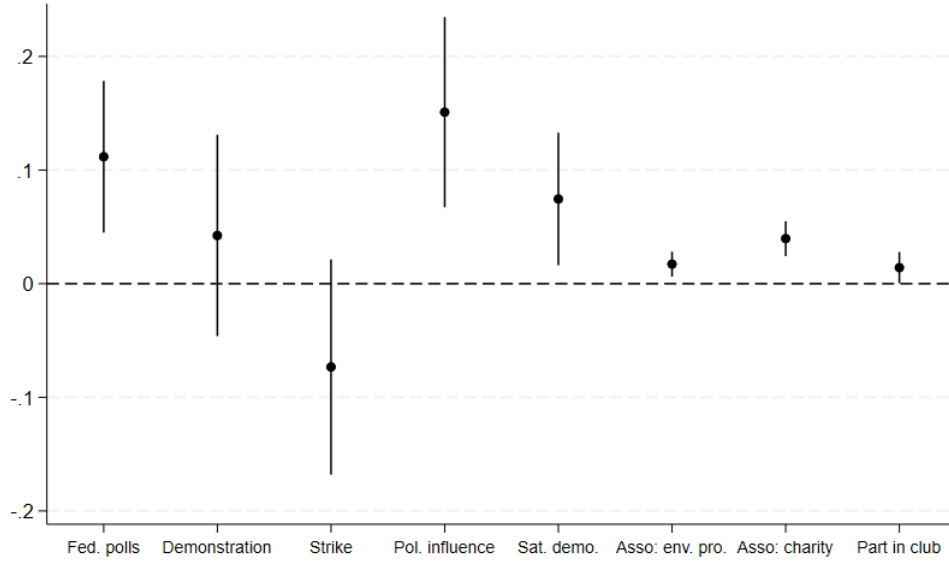
Notes: The figure replicates the specification in Equation (6), reporting the coefficient on $UNIV_{ict}$. Full sample is considered (Swiss individuals aged 18 and over). Reporting 90% confidence intervals based on standard errors clustered at the individual level.

5 Behind the State Normative Effect: State Normative Legitimacy

The model formalizes that the effect of the state's normative composition on individual deliberative normative transformation is not driven by the personal universalist values of state agents, but by the symbolic authority of the social norm they are institutionally mandated to uphold. This norm grounds their moral legitimacy as representatives of the state. To identify this causal mechanism, I leverage the workplace setting and examine dynamic changes in universalism endorsement among "switchers": individuals who switch between the private and public sectors – outside of public service (mission-oriented) industries. Specifically, I test whether working for the state causally increases universalism endorsement through the deliberative recognition that the social norm underpins institutional legitimacy[12].

---

[12]Organizational psychologists have long emphasized the importance of workplace socialization in shaping cultural change (Schein (1965)). According to Van Maanen and Schein (1979), workplace

As I adopt a within-individual estimation strategy to identify the causal effect of working for the state on norm endorsement, the source of identification comes from the switchers. I sample individuals aged 20-54, to focus on voluntary job-to-job mobility (Groot and Verberne (1997)), who are employed in the public or private sector (excluding the self-employed). The final sample contains 3'777 individuals (9'849 person observations), of which 2'369 work outside public service industries. Among these, 8,6% switch institutional sectors during the sample period (1999-2003); transitions occur equally in both directions, and no individual switches more than once. Institutional switches are common in Switzerland, due to the high permeability between public and private employment. This permeability stems from open hiring processes and the absence of a formal civil service status[13]. Importantly, public sector jobs offer a wage premium: on average 12%, and up to 29% at the lower end of the pay scale (Portmann et al. (2024)).

5.1 Empirical Strategy

I estimate the following fixed effects regression where the state effect is identified only from individuals who change institutional sector:

$$UNIV_{ict} = \eta \times State_{it} + \psi \times Fed_{it} + X'_{ct}\delta + \gamma_i + \epsilon_{ict}. \tag{6}$$

The unit of observation is a worker (i) $\times$ canton (c) $\times$ year (t) cell. The variable $State$ is equal to 1 if the individual works for the state (at the cantonal or federal level) and 0 if the individual works in the private (both for-profit and nonprofit) sector. Fed is equal to 1 if the individual works for the state at the federal level and 0 otherwise. I include individual fixed effects γ_i. X_{ct} is a vector of canton- specific year dummies to account for time-varying confounders that operate at the cantonal level (e.g. economic distress driving the odds of working in the public sector and individual preferences). As before, the variable $UNIV$ is equal to 1 if the individual endorses the universalism social norm and 0 otherwise. Standard errors are clustered at the individual level.

The within estimator strongly alleviates the selection into occupation concern. However, if workers choose the time spent in an institutional sector as a function of their own dynamics in preferences, then workers who experience a shock to their social norm such that they are aligned with the state's normative legitimacy might switch into the public sector. In the next section, I provide evidence on who the switchers are.

5.2 Who are the Switchers?

5.2.1 Addressing the Dynamic Self-selection Concern

The existing theoretical (Francois (2003)) and empirical (Gregg et al. (2011)) literature formalizes and finds that agents who have a civic-minded interest in service self-select

---

socialization or "organizational socialization refers...to the fashion in which an individual is taught and learns what behaviors and perspectives are customary and desirable within the work setting as well as what ones are not" (Van Maanen and Schein (1979), p. 4).

[13]The Swiss public sector largely operates under an "open" civil service system, meaning no competitive exam is required (Audier and Bacache-Beauvallet (2007)). Consequently, worker profiles in the public sector are more heterogeneous. Emery et al. (2014) show that individuals with private sector experience are frequent at all hierarchical levels. Among a sample of 1,045 public-sector top executives, Guido Schilling AG (2023) finds that one-third have private sector experience. Moreover, since the 1990s, the formal civil servant status has been abolished in all Swiss cantons except two.

into caring services (defined as health, education and social care), which are delivered by both the public and private sectors but more likely by the public sector. This implies that self-selection based on the universalism social norm is at the service-type level rather than at the institutional level.

To test this selection hypothesis, I investigate pre-trends in the social norm of switchers – comparing their initial social norm (before switching) with that of stayers in the same sector. I consider four binary indicators representing the public and private non-HES sectors and the public and private HES sectors and estimate the following models:

$$UNIV_{ict}^{Private} = \phi \times Switch_i^{State} + X_{ct}'\delta + Z_{it}'\gamma + u_{it} \tag{7}$$

$$UNIV_{ict}^{PrivateNonHES} = \phi \times Switch_i^{StateHES} + X_{ct}'\delta + Z_{it}'\gamma + u_{ict} \tag{8}$$

$$UNIV_{ict}^{PrivateHES} = \phi \times Switch_i^{StateNonHES} + X_{ct}'\delta + Z_{it}'\gamma + u_{ict} \tag{9}$$

$$UNIV_{ict}^{PrivateHES} = \phi \times Switch_i^{StateHES} + X_{ct}'\delta + Z_{it}'\gamma + u_{ict} \tag{10}$$

$$UNIV_{ict}^{PrivateNonHES} = \phi \times Switch_i^{StateNonHES} + X_{ct}'\delta + Z_{it}'\gamma + u_{ict} \tag{11}$$

where the variable $UNIV_{ict}^{Private}$ is equal to 1 if the individual endorses the universalism social norm and 0 otherwise when he works in the private sector (or in other types of private sector). $Switch_i^{State}$ is an indicator variable equal to one if the individual switches into the public sector (or into specific industries in the public sector) at any point in the future. X_{ct} is a vector of canton-specific year fixed effects to account for time-varying confounders that operate at the cantonal level. I also include a set of control variables for individual characteristics (gender, number of children, age, religion, culture, college education, marital status, health status, urban) in Z_{it}. Standard errors are clustered at the individual level. $StateHES$ stands for occupations in health, education and social care services in the public sector and $StateNonHES$ all other occupations in the public sector (and similarly for the $PrivateHES$ and $PrivateNonHES$ variables in the private sector). The coefficient on the switching indicator therefore captures systematic differences in universalism between those who stay in the private sector and those who switch out into the public sector at some future point. In addition, to allow for different selection patterns across institutional levels, I run the same regressions while excluding the federal level.

Figure 10 in Appendix A summarizes regression coefficients of the switching indicators in all equations. They confirm that self-selection based on universalism is at the service-type level rather than at the institutional sector level. Specifically, I find no evidence of workers' selection into the public sector based on their preference for the universalism social norm within non-HES occupations as well as within HES occupations. Furthermore, there is no evidence of a different selection pattern across institutional levels within the public sector.

5.2.2 Changing Selection with Switchers

I also examine changing selection in the public sector with switchers. Results are presented in Table 10 in Appendix A.3. Compared to workers who initially chose the public non-HES sector, those switching from the private non-HES sector to the public non-HES sector are significantly younger, though by less than two years on average, with a mean age of 39.

While experiences during the so-called "formative years" can shape later-life preferences and views (Roth and Wohlfart (2018), Aneja et al. (2024)), switchers are evidently beyond this stage.

Switchers are also significantly less likely to be college educated. Interestingly, Ashraf et al. (2020) find a trade-off in sorting between mission and financial motivation for low-level talent. As highlighted, in Switzerland, there is a notable public sector wage premium for similar jobs with equivalent qualifications and hierarchical status. While on average, the public sector wage premium is 12%, it can reach up to 29% at the lower end of the pay scale (Portmann et al. (2024)). Therefore, it is highly likely that the switching bureaucrats in my sample are driven by financial motives.

5.3 Working for the State Induces a Within-individual Shift in the Universalism Norm Only

Table 5 reports the results of estimating equation (7), using a fixed-effects linear probability model. Results are robust to using a random-effects logit model (available upon request). Column (1) identifies the effect of working for the state (η) using all switchers between the private and public sectors. The estimated coefficient is positive and statistically significant, suggesting that state employment increases the likelihood of endorsing the universalism norm. Column (2) restricts the sample to switches occurring within the same industry type (non-HES and HES) to address selection into occupation. Column (3) focuses on switches from the private to the public sector, and Column (4) further narrows the sample to private-to-public switches within non-HES industries – this serves as the baseline estimate. Across these specifications, the estimated effect grows stronger, as expected. In Column (4), working for the cantonal state increases the probability of endorsing the universalism norm by 10 percentage points relative to when the same individual is employed in a similar job in the private sector. This represents a sizable effect – more than 15% relative increase, given that the mean level of universalism endorsement in the private non-HES sector is 65%. Figure 11 in Appendix A.3 confirms that results are not driven by specific cantons.

By contrast, the negative and statistically significant coefficient for federal bureaucracies indicates that working at this institutional level has no global effect on norm endorsement. This suggests that the symbolic norm of universalism is not embedded within the federal institutional setting. This finding is consistent with earlier evidence, particularly the 11 percentage-point gap in universalism endorsement between cantonal and federal state agents, and resonates with Ekeh (1975)'s concept of the two publics in the African context: one local, with moral legitimacy, and the other central and amoral.

Column (5) presents results for switches from the public to the private sector. The estimated coefficient is negative but statistically insignificant. However, due to the smaller number of observations in this group, the power of the test is limited. Therefore, it is difficult to draw firm conclusions about whether the private sector erodes the universalism social norm or not.

In Table 6, I replicate the specification from Column (4) of Table 5 using the alternative societal preference outcomes previously examined in Table 4. The results reveal no evidence that working for the state significantly alters bureaucrats' preferences regarding welfare, redistribution, defense spending, environmental protection, political ideology, or trust in the federal government. However, it does appear to make them more critical of the quality of democracy. This could reflect heightened normative expectations induced

Table 5: Working for the state makes individuals endorse the universalism social norm

UNIV	(1) All switches	(2) All switches within non-HES & HES	(3) Private to public within non-HES & HES	(4) Private to public within non-HES	(5) Public to private within non-HES
State	0.0458**	0.0526**	0.0838**	0.0987**	
	(0.0202)	(0.0266)	(0.0340)	(0.0488)	
State: Federal	-0.0512*	-0.0477	-0.114**	-0.122*	
	(0.0271)	(0.0348)	(0.0549)	(0.0687)	
Private sector					-0.0274
					(0.0427)
Observations	8,672	4,709	6,343	4,531	1,084
Number of individuals	3,590	2,222	2,794	1,960	441

Notes: All regressions include individual fixed effects as well as canton-specific year fixed effects. Robust standard errors are clustered at the individual level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

by institutional exposure.

Table 6: Working for the state does not affect other societal preferences

	(1) Welfare	(2) Redistri.	(3) Environment	(4) Army	(5) Pol. ideology	(6) Trust fed. gov.	(7) Sat. with democracy
State	-0.105	0.0289	-0.0357	-0.0486	-0.0905	-0.300	-0.380*
	(0.107)	(0.122)	(0.0803)	(0.115)	(0.209)	(0.263)	(0.226)
State: Federal	0.167	-0.0518	0.183	0.127	0.117	0.234	0.191
	(0.171)	(0.153)	(0.116)	(0.139)	(0.337)	(0.326)	(0.311)
Observations	4,980	5,036	2,988	5,031	4,451	5,091	5,046
Number of individuals	2,030	2,040	1,559	2,036	1,898	2,052	2,039

Notes: Sample restricted to switches across institutional sectors within non-HES industries. All regressions individual individual fixed effects as well as canton-specific year fixed effects. Robust standard errors are clustered at the individual level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

5.4 Amplification Through Cultural Variation in State Normative Legitimacy

Since public bureaucrats may endorse universalism without internalizing it – due to the public sector's specific institutional setup of incentives and constraints – I first investigate whether the estimated causal effect of working for the state is mediated by these external factors. Column (1) of Table 7 shows that the public sector effect increases rather than attenuates compared to the baseline counterpart, and the added controls have coefficients close to zero. This pattern suggests that the institutional setup does not mediate the relationship between state employment and universalism endorsement.

Next, I explore heterogeneity in effect size. Column (2) replicates the specification from Column (3) in Table 5, allowing the state effect to vary by sector – specifically, between caring and non-caring professions (HES vs. non-HES). The results show no significant difference across sectors. This suggests that the observed effect is not driven by spillovers from the organizational culture of mission-oriented jobs, but instead reflects the symbolic institutional role of the state itself. In other words, what matters is where individuals work – the public sector as a locus of institutional legitimacy – not what

they do within it. This finding is consistent with a core assumption of my model: that normative legitimacy is attached to institutions broadly.

In Column (3), I examine cultural heterogeneity in the effect of state employment. French-speaking Swiss individuals are more exposed to French political culture and, in particular, to the state's normative motivation as theorized in the French tradition. Rousseau's political philosophy was the dominant intellectual force at the onset of the French Revolution (Bensel (2022)). In his conception of the General Will, the state metaphorically embodies collective moral purpose[14]. Under this tradition, the state is not merely functional but normatively charged, and this symbolic role has been codified institutionally: the French Constitution of October 27, 1946, still valid today, asserts in its preamble (paragraph 9): "All property and all enterprises that have or that may acquire the character of a public service or de facto monopoly shall become the property of society"[15]. In the same vein, Weber (1976) argues that civil servants in the French tradition have long served not merely as bureaucrats or law enforcers, but as normative embodiments of the state's presence in local life. Through their daily presence in local bureaucracies, they came to represent the symbolic authority of the Republic and played a central role in promoting its universalist ideals among citizens – particularly in peripheral or initially resistant communities, in the spirit of (*Peasants into Frenchmen*). This cultural tradition may foster stronger endorsement of the universalism norm among French-speaking Swiss, who are more likely to view the state as a legitimate moral institution in their own country – in other words, α in the theoretical model is higher for them. This interpretation is supported by the individual-level correlates of universalism endorsement shown in Figure 9 in Appendix A.1 as well as by the greater convergence – lower normative dispersion – toward the social norm revealed in Figure 7.

Accordingly, I expect the effect of working for the state on universalism endorsement to be more pronounced among French-speaking Swiss bureaucrats, reflecting greater treatment intensity despite their already higher baseline levels of universalism endorsement. The empirical specification includes individual fixed effects and canton-by-year fixed effects, ensuring that the interaction term between state employment and French-speaking identity identifies the within-individual differential response to state employment across linguistic groups, net of time-invariant individual traits and regional institutional environments, knowing that I find no evidence of different patterns of selection into the public sector between French-speaking and German-speaking Swiss individuals based on their universalism. Consistent with this prediction, the estimated effect nearly doubles for them, reaching approximately 20 percentage points, suggesting that the effect of working for the state operates through deliberative recognition of the state's symbolic legitimacy.

To illustrate this effect, I draw on William H. Sewell (2005)'s example of a worker who, convinced by the marginal productivity theory of distribution, believes he is earning the value of his marginal product – and this is a just outcome or at least it is consistent with the First Welfare Theorem. During lunch, he reads *The Communist Manifesto* by Marx

---

[14]Bensel (2022), p. 208, notes: "Depicting modern society as both corrupt and depraved, Rousseau thought that the only way to recover innocence was to empower the General Will of the people, because, when properly elicited, the General Will unfailingly reflected the universal good of society. Under the social contract thus formed: 'Each of us puts his goods, his person, his life, and all his power in common under the supreme direction of the general will, and we as a body accept each member as a part indivisible from the whole.' (from *Émile*. McPhee, *Liberty or Death*, p.31)".

[15]This linkage between the state and universal goods explains why they are referred to as *public services* in France and must be publicly owned. By contrast, the European Union – under German influence – uses the term *services of general interest* to avoid this public-sector linkage.

and Engels, leading to a paradigm shift – now believing he is being exploited – and returns to his job as a completely different person, even though the job itself remains unchanged. Similarly, consider a worker – a cook – who switches from the private to the public sector for better pay, while performing the same job. In the morning, he prepares meals in the cafeteria of the Department of Finance, still holding the exclusionary social norm that Switzerland should prioritize Swiss citizens over foreigners. However, during lunchtime conversations with his new co-workers, discussions emerge about the normative legitimacy the state is expected to uphold – a legitimacy they represent, even when it conflicts with their private values. Through these interactions, he is introduced to Castoriadis' *The Imaginary Institution of Society*. When he returns to his duties, the physical environment, technology, and production process remain unchanged. Yet, his paradigm has shifted: as a state agent, he now believes he should endorse the universalism social norm.

Figure 7: Normative dispersion by linguistic region over time
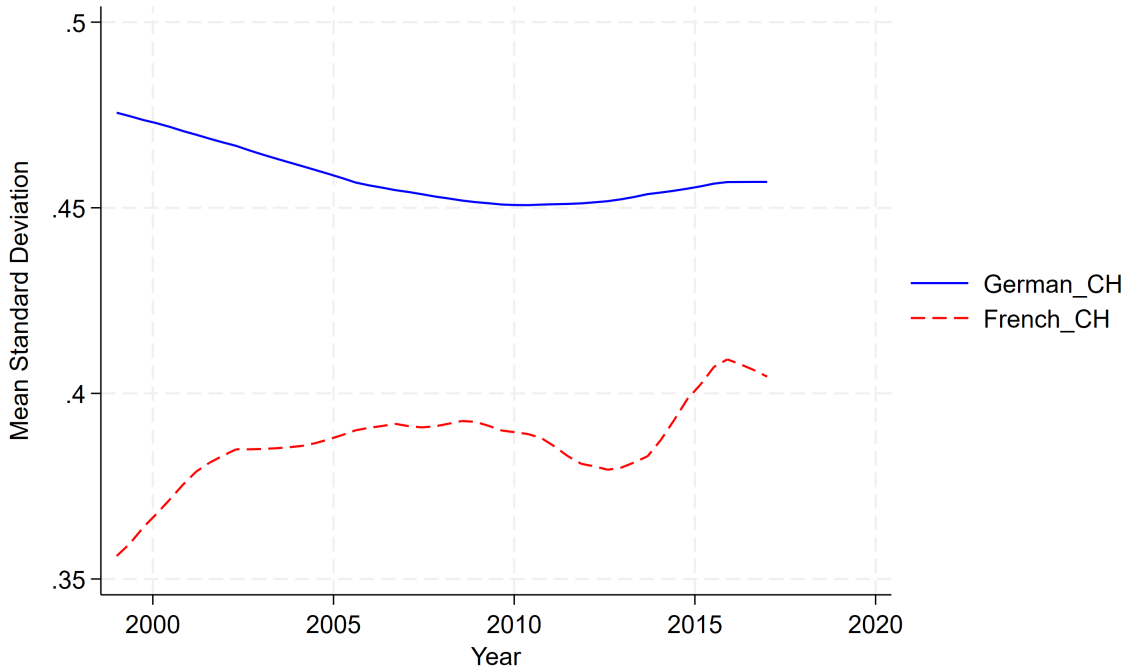
Note: The trend lines depict LOESS-smoothed curves tracing the evolution of normative dispersion over time within each linguistic region (German-speaking and French-speaking cantons, excluding bilingual ones). Normative dispersion is measured as the average within-canton standard deviation of individual responses to the universalism social norm question among Swiss individuals aged 18 and over, computed separately for each region and year.

As robustness check to ensure that the French-speaking effect is not simply capturing other cultural differences – such as political ideology – than increased state normative legitimacy, I draw on the findings of Eugster and Parchet (2019). Using data from federal referendum outcomes at the municipality level between 1981 and 2011, they find that voters in French-speaking municipalities consistently show significantly stronger support for policies recommended by left-of-center parties, with a discrete jump of approximately 8 percentage points at the language border. To verify that the stronger public sector effect among French-speaking Swiss is not merely a reflection of this left-leaning political

orientation, I re-estimate the model allowing the effect of public sector employment to vary by individuals' own political ideology, using the same ideological categories as in Figure 4. The results, reported in Column (4), reveal that the effect is in fact weaker among left-wing individuals. This pattern is inconsistent with the French-speaking result being driven by political ideology, and instead reinforces the interpretation that the cultural effect is rooted in a stronger normative conception of the state – consistent with the model's α parameter.

Table 7: What drives the effect of working for the state?

	(1)	(2)	(3)	(4)
		Universalism Endorsement		
State	0.103*	0.0967**	0.0671	0.178**
	(0.0561)	(0.0450)	(0.0548)	(0.0748)
State: Federal	-0.133*	-0.127**	-0.144**	-0.141**
	(0.0780)	(0.0604)	(0.0695)	(0.0688)
Ln yearly income	0.00202			
	(0.0250)			
Sat. with work colleagues	-0.00332			
	(0.00587)			
Job Security	0.00454			
	(0.0106)			
State x HES		-0.0201		
		(0.0602)		
State x French-Swiss			0.127**	
			(0.0631)	
State x Left				-0.195***
				(0.0738)
Observations	3,960	5,338	4,531	3,947
Number of individuals	1,805	2,316	1,960	1,801

Notes: The sample includes switches only from private non-HES to public non-HES sector in Columns (1) and (3-4). Column (2) considers switches within both non-HES and HES industries. All regressions include individual fixed effects as well ass canton-specific year fixed effects. Robust standard errors are clustered at the individual level. *$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

Overall, these results suggest that the state effect captured in Equation (5) is a "state normative legitimacy" effect, whereby public employees embody the universalism social norm as part of their institutional role.

6 Conclusion

I developed a model of autonomy in norm formation. Whereas social norms are usually thought of as emerging from the aggregation of fixed private social preferences, I conceptualize them as endogenously created through costly normative effort. Rational individuals exert this effort because norms legitimize institutional functionality. In autonomous societies, therefore, individuals actively self-institute both the social norms and the formal institutions that embody them, leading to a co-evolution of normative orders and institutional structures. Crucially, autonomy is a collective deliberative process: individual deliberation over norms does not occur in a vacuum but within a deliberative environment shaped by both state and peer normative embodiment. At the core of autonomy lies layered agency – the capacity for critical moral reasoning not only about oneself but also about one's society, allowing for deliberate normative or strategic divergences between

personal and societal stances. Autonomy thus offers a richer account of individual agency than the heteronomous perspective, which treats individuals as shaped by institutions but not as their co-creators. Autonomy implies that stable social norms are not fixed constraints but dynamic equilibria, open to endogenous change, including rapid shifts, as radical imaginaries emerge and diffuse through the public sphere.

This conception of autonomy offers a renewed normative foundation for democratic and collective engagement. When individuals are aligned with the social norm and see themselves as its ongoing co-creators, they come to view democratic participation not merely as a means of defending private interests but an expression of shared responsibility in shaping the normative order that underpins institutional legitimacy. Democracy, then, is not merely a mechanism for aggregating preferences; it is the institutional expression of an autonomous society capable of norm self-institution. Consequently, the state is not simply a functional apparatus, but the embodiment of the social norm. Accordingly, all state agents, including bureaucrats, play a central role in the dynamics of the instituted normative order.

Using individual-level, geolocated panel data from Switzerland on endorsement of the universalism social norm, the empirical analysis provides strong support for the the model's autonomy-based assumptions and predictions. The findings underscore the enduring influence of the Western European Enlightenment tradition, marked by rationality, critical reflection, and the imperative that all aspects of social life, including the moral order, be justified without recourse to divine or natural authority. Crucially, however, autonomy is not a fixed condition. It can erode under authoritarian drift or though a loss of societal capacity to collectively reflect on and reshape normative orders.

References

Abeles, Marc (1988) "Modern Political Ritual: Ethnography of an Inauguration and a Pilgrimage by President Mitterrand," *Current Anthropology*, 29 (3), 391–404.

Acemoglu, Daron, Davide Cantoni, Simon Johnson, and James A. Robinson (2011) "The Consequences of Radical Reform: The French Revolution," *American Economic Review*, 101 (7), 3286–3307, 10.1257/aer.101.7.3286.

Acemoglu, Daron and Matthew O. Jackson (2015) "History, Expectations, and Leadership in the Evolution of Social Norms," *The Review of Economic Studies*, 82 (2), 423–456, 10.1093/restud/rdu039.

——— (2017) "Social Norms and the Enforcement of Laws," *Journal of the European Economic Association*, 15 (2), 245–295, 10.1093/jeea/jvw006.

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001) "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review*, 91 (5), 1369–1401, 10.1257/aer.91.5.1369.

Acemoglu, Daron and James A. Robinson (2012) *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*, New York: Crown Publishing Group.

——— (2025) "Culture, Institutions, and Social Equilibria: A Framework," *Journal of Economic Literature*, Forthcoming.

Anderson, Benedict (1983) *Imagined Communities: Reflections on the Origin and Spread of Nationalism*, London: Verso.

Aneja, Abhay, Farina Silvia, and Xu Guo (2024) "Beyond the War: Public Service and the Transmission of Gender Norms," working paper.

Arendt, Hannah (1958) *The Human Condition*, Chicago: University of Chicago Press.

——— (1963) *Eichmann in Jerusalem: A Report on the Banality of Evil*, New York: The Viking Press.

Ashraf, Nava, Oriana Bandiera, Edward Davenport, and Scott S. Lee (2020) "Losing Prosociality in the Quest for Talent? Sorting, Selection, and Productivity in the Delivery of Public Services," *American Economic Review*, 110 (5), 1355–94, 10.1257/aer.20180326.

Athias, Laure and Bruno Ventelou (2025) "Universalism Across Allocation Channels: Evidence of Normative Layering Within Individuals," working paper.

Athias, Laure and Pascal Wicht (2025) "Make or buy for public services: Culture matters for efficiency considerations," *The Quarterly Review of Economics and Finance*, 99, 101899, https://doi.org/10.1016/j.qref.2024.101899.

Audier, F. and M. Bacache-Beauvallet (2007) "Emploi dans la fonction publique et fonctions "d'intérêt public": Que nous apprennent les comparaisons internationales ?" *Revue de l'OFCE*, 103, 323–350, 10.3917/reof.103.0323.

Becker, Sascha O. and Hans-Joachim Voth (2025) "From the Death of God to the Rise of Hitler," working paper.

Bensel, Richard Franklin (2022) *The Founding of Modern States*: Cambridge University Press, doi.org/10.1017/9781009247245.

Besley, Timothy, Robin Burgess, Adnan Khan, and Guo Xu (2022) "Bureaucracy and Development," *Annual Review of Economics*, 14 (Volume 14, 2022), 397–424, https://doi.org/10.1146/annurev-economics-080521-011950.

Besley, Timothy and Maitreesh Ghatak (2005) "Competition and Incentives with Motivated Agents," *American Economic Review*, 95 (3), 616–636, 10.1257/0002828054201413.

Bisin, Alberto and Thierry Verdier (2001) "The economics of cultural transmission and the dynamics of preferences," *Journal of Economic theory*, 97 (2), 298–319.

Blanc, Guillaume and Masahiro Kubo (2023) *State-sponsored education and French identity*, London: CEPR Press.

Bond, R., C. Fariss, J. Jones, A. D. I. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler (2012) "A 61-million-person experiment in social influence and political mobilization," *Nature*, 489, 295–298, 10.1038/nature11421.

Bondar, Mariia and Nicola Fuchs-Schündeln (2023) "Good Bye Lenin Revisited: East-West Preferences Three Decades after German Reunification," *German Economic Review*, 24 (1), 97–119, doi:10.1515/ger-2022-0042.

Bourdieu, Pierre (1994) *Raisons Pratiques: Sur la Théorie de L'Action*, Paris: Seuil.

Bursztyn, Leonardo, Georgy Egorov, and Stefano Fiorin (2020) "From Extreme to Mainstream: The Erosion of Social Norms," *American Economic Review*, 110 (11), 3522–48, 10.1257/aer.20171175.

Bénabou, Roland, Davide Ticchi, and Andrea Vindigni (2021) "Forbidden Fruits: The Political Economy of Science, Religion, and Growth," *The Review of Economic Studies*, 89 (4), 1785–1832, 10.1093/restud/rdab069.

Bénabou, Roland and Jean Tirole (2003) "Intrinsic and Extrinsic Motivation," *The Review of Economic Studies*, 70 (3), 489–520, https://doi.org/10.1111/1467-937X.00253.

——— (2011) " Identity, Morals, and Taboos: Beliefs as Assets," *The Quarterly Journal of Economics*, 126 (2), 805–855, 10.1093/qje/qjr002.

——— (2025) " Laws and Norms," *Journal of Political Economy*, Forthcoming.

Cappelen, Alexander W., Benjamin Enke, and Bertil Tungodden (2025) "Universalism: Global Evidence," *American Economic Review*, 115 (1), 43–76, 10.1257/aer.20230038.

Cassar, Lea and Stephan Meier (2018) "Nonmonetary Incentives and the Implications of Work as a Source of Meaning," *Journal of Economic Perspectives*, 32 (3), 215–38, 10.1257/jep.32.3.215.

Castoriadis, Cornelius (1987) *The Imaginary Institution of Society*, Cambridge, Massachusetts: The MIT Press.

——— (1991) *Philosophy, politics, autonomy. Essays in political philosophy*, New York: Oxford University Press.

——— (1996) *La montée de l'insignifiance: Les carrefours du labyrinthe IV*, Paris: Éditions du Seuil.

——— (1997) *World in fragments. Writings on politics, society, psychoanalysis and the imagination*, Stanford: Stanford University Press.

Clots-Figueras, Irma and Paolo Masella (2013) "Education, Language and Identity," *The Economic Journal*, 123 (570), 10.1111/ecoj.12051.

Ekeh, Peter P. (1975) "Colonialism and the Two Publics in Africa: A Theoretical Statement," *Comparative Studies in Society and History*, 17 (1), 91–112.

Emery, Y., D. Giauque, and F. Rebmann (2014) "La lente mutation des élites administratives fédérales suisses," *Revue Internationale des Sciences Administratives*, 80, 725–747, 10.3917/risa.804.0725.

Enke, Benjamin, Ricardo Rodriguez-Padilla, and Florian Zimmermann (2022) "Moral Universalism: Measurement and Economic Relevance," *Management Science*, 68 (5), 3175–3973, https://doi.org/10.1287/mnsc.2021.4086.

——— (2023) "Moral Universalism and the Structure of Ideology," *The Review of Economic Studies*, 90 (4), 1934–1962, 10.1093/restud/rdac066.

Eugster, Beatrix and Raphaël Parchet (2019) "Culture and Taxes," *Journal of Political Economy*, 127 (1), 296–337, 10.1086/700760.

Francois, Patrick (2003) "Not-for-Profit Provision of Public Services," *The Economic Journal*, 113 (486), C53–C61.

Geertz, Clifford (1973) *The Interpretation of Cultures*, New York: Basic Books.

Gregg, Paul, Paul Grout, Anita Ratcliffe, Sarah Smith, and Frank Windmeijer (2011) "How important is pro-social behaviour in the delivery of public services?" *Journal of Public Economics*, 95 (7), 758–766, `https://EconPapers.repec.org/RePEc:eee:pubeco:v:95:y:2011:i:7:p:758-766`.

Groot, Wim and Maartje Verberne (1997) "Aging, Job Mobility, and Compensation," *Oxford Economic Papers*, 49 (3), 380–403.

Guido Schilling AG (2023) "Schilling report 2023: Transparency at the Top - The Management Boards of Switzerland's Private and Public Sector,"Technical report.

Habermas, Jürgen (1989) *The Structural Transformation of the Public Sphere: An Inquiry into a Category of Bourgeois Society*: MIT Press.

Harsanyi, John C. (1955) "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, 63 (4), 309–321.

Heldring, Leander (2020) "The Origins of Violence in Rwanda," *The Review of Economic Studies*, 88 (2), 730–763, 10.1093/restud/rdaa028.

——— (2023) "Bureaucracy as a tool for Politicians: Evidence from Germany," working paper.

Henrich, Joseph (2020) *The WEIRDest People in the World: How the West Became Psychologically Peculiar and Particularly Prosperous*, New York: Farrar, Strauss, and Giroux.

Hobsbawm, E. J. (1990) *Nations and Nationalism since 1780: Programme, Myth, Reality*, Cambridge: Cambridge UniversityPress.

Jones, J. J., R. Bond, E. Bakshy, D. Eckles, and J. H. Fowler (2017) "Social influence and political mobilization: Further evidence from a randomized experiment in the 2012 U.S. presidential election," *PloS one*, 12 (4), 10.1371/journal.pone.0173851.

Kohn, Hans (1950) "Romanticism and the Rise of German Nationalism," *The Review of Politics*, 12 (4), 443–472.

Kranton, Rachel E. (2019) "The Devil Is in the Details: Implications of Samuel Bowles's *The Moral Economy* for Economics and Policy Research," *Journal of Economic Literature*, 57 (1), 147–160, https://doi.org/10.1257/jel.20171463.

Lazarsfeld, Berelson B., P. F. and H. Gaudet (1944) *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign*, New York: Duell, Sloan & Pearce.

Levi, Margaret (1997) *Consent, Dissent, and Patriotism*, Cambridge: Cambridge University Press.

Mokyr, Joel (2012) *Culture, institutions, and modern growth*, 151–191: Cambridge University Press, 10.1017/CBO9781107300361.010.

———— (2016) *A Culture of Growth: The Origins of the Modern Economy*, Princeton, NJ: Princeton University Press.

Mosse, George L. (1964) *The Crisis of German Ideology: Intellectual Origins of the Third Reich*, New York: Grosset and Dunlap.

Ostrom, Elinor (1990) *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge: Cambridge University Press.

PewResearchCenter (2011) "The American-Western European Values Gap,"Technical report, Pew Research Center's Global Attitudes Project, Retrieved from http://pewrsr.ch/vIHVsk.

Portmann, Marco, Frederik Blümel, and Christoph A. Schaltegger (2024) "Lohnprämie in den öffentlichen Verwaltungen: Neue Analyse 2024," Policy Papers 18, IWP.

Putnam, Robert D., Robert Leonardi, and Raffaella Y. Nonetti (1993) *Making Democracy Work: Civic Traditions in Modern Italy*: Princeton University Press.

Rawls, John (1971) *A Theory of Justice*, Cambridge, MA: Harvard University Press.

Roth, Christopher and Johannes Wohlfart (2018) "Experienced inequality and preferences for redistribution," *Journal of Public Economics*, 167, 251–262, https://doi.org/10.1016/j.jpubeco.2018.09.012.

Sahlins, Marshall (1976) *Culture and Practical Reason*, Chicago: The University of Chicago Press.

———— (2022) *The New Science of the Enchanted Universe: An Anthropology of Most of Humanity*, Princeton: Princeton University Press.

Sandel, Michael J. (1998) *Liberalism and the Limits of Justice*, Cambridge, MA: Cambridge University Press, 2nd edition.

Satyanath, Shanker, Nico Voigtländer, and Hans-Joachim Voth (2017) "Bowling for Fascism: Social Capital and the Rise of the Nazi Party," *Journal of Political Economy*, 125 (2), 478–526, 10.1086/690949.

Schein, Edgar H. (1965) *Organizational Psychology*, Englewood Cliffs, NJ: Prentice-Hall.

Shiller, Robert J. (2017) "Narrative Economics," *American Economic Review*, 107 (4), 967–1004, 10.1257/aer.107.4.967.

Tabellini, Guido (2008a) "The Scope of Cooperation: Values and Incentives," *The Quarterly Journal of Economics*, 123 (3), 905–950.

———— (2008b) "Institutions and Culture," *Journal of the European Economic Association*, 6 (2-3), 255–294, 10.1162/JEEA.2008.6.2-3.255.

Taylor, Charles (2004) *Modern Social Imaginaries*, Durham and London: Duke University Press.

Tilly, C. (1975) *Reflections on the History of European State-Making*, Princeton: Princeton University Press.

Tirole, Jean (2023) "Safe Spaces: Shelters or Tribes?," working paper.

Tyler, Tom R. (2006) *Why People Obey the Law*: Princeton University Press.

Van Maanen, John E. and Edgar H. Schein (1979) "Toward a theory of organizational socialization," *Research in Organizational Behavior*, 1, 209–269.

Walter, André and Patrick Emmenegger (2019) "The Partisan Composition of Cantonal Governments in Switzerland, 1848-2017. A New Data Set," *Swiss Political Science Review*, 25 (1), 1–18, https://doi.org/10.1111/spsr.12329.

Weber, Eugen (1976) *Peasants into Frenchmen: The Modernization of Rural France, 1870-1914*, Stanford, Calif.: Stanford University Press.

Weber, Max (1922) *Economy and Society*, 4th ed. Tübingen.

William H. Sewell, Jr (2005) *Logics of History: Social Theory and Social Transformation*, Chicago: University of Chicago Press.

A   Appendix: Figures and Tables

A.1   Descriptive statistics

Figure 8: Endorsement of the universalism social norm over time
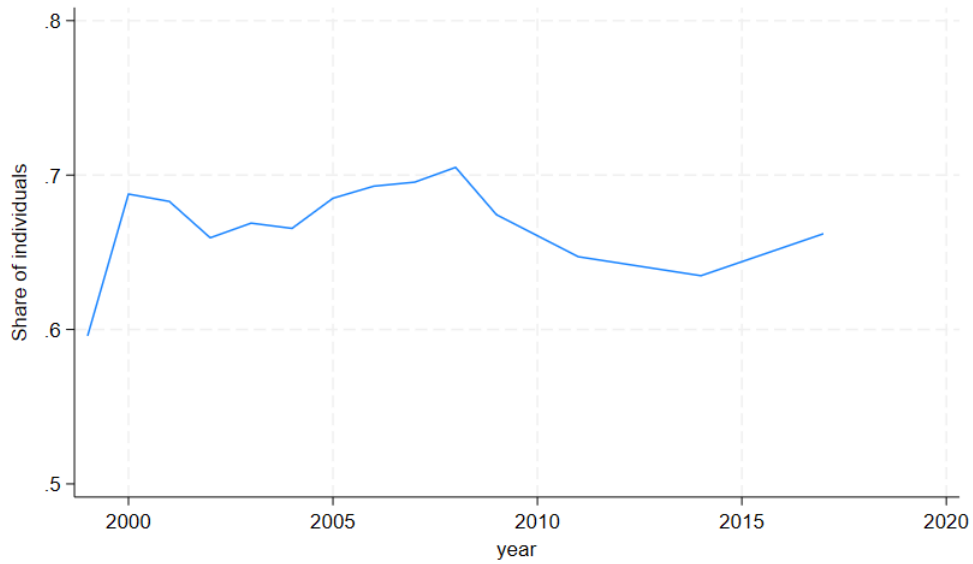
Note: The figure depicts the share of Swiss individuals aged 18 and over who endorse the universalism social norm over the maximal period of time with data availability, corresponding to the yearly average of the variable $UNIV$.

Figure 9: Correlates of endorsement of the universalism social norm
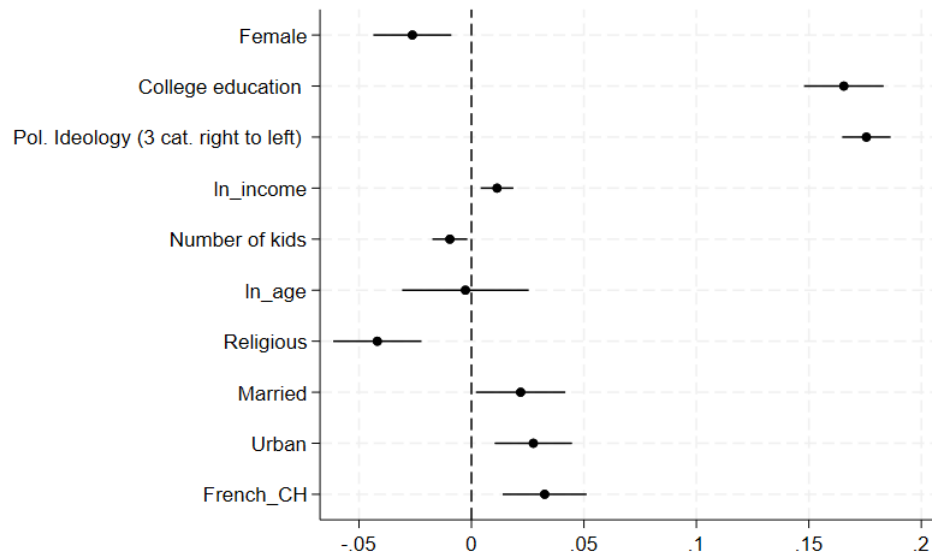
Note: This figure presents OLS Estimates from regressions of the $UNIV$ variable on all sociodemographic correlates, with SE clustered at the individual level, along with 95% confidence intervals.

A.2 Baseline: Robustness Checks

Table 8: Random-effects logit model, Effect of State and Peer Norms on Universalism Endorsement among Non-Workers

	(1)	(3)	(5)
	Individual Universalism Endorsement		
Share UNIV among state employees	1.602**	1.622**	1.515**
	(0.682)	(0.687)	(0.686)
Left-wing state		-0.0636	-0.0881
		(0.234)	(0.235)
Share UNIV among peers			1.274*
			(0.696)
Observations	10,841	10,841	10,841
Number of individuals	4,575	4,575	4,575

Notes: All regressions are estimated using a random-effects logit model and include canton and year fixed effects. The binary dependent variable equals 1 if the non-working individual endorses the universalism social norm, and 0 if they favor preferential treatment for Swiss citizens. *Share UNIV among state employees* and *Share UNIV among peers* refer to canton-year averages of universalism endorsement among state and private sector employees, respectively. *Left-wing state* is a binary indicator equal to 1 if the cantonal executive is left-leaning (see text for coding details). Standard errors are computed using the robust (Huber–White) estimator. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Table 9: $UNIV$ as a three-point ordinal variable, Effect of State and Peer Norms on Universalism Endorsement among Non-Workers

	(1)	(2)	(3)
	Individual Universalism Endorsement		
Share UNIV among state employees	0.286**	0.281**	0.258*
	(0.139)	(0.140)	(0.140)
Left-wing state		0.0135	0.00711
		(0.0481)	(0.0483)
Share UNIV among peers			0.330**
			(0.145)
Observations	12,100	12,100	12,099
Number of individuals	4,779	4,779	4,778

Notes: All regressions are estimated using ordinary least squares (OLS) and include individual fixed effects, as well as canton and year fixed effects. The dependent variable is a three-point ordinal measure coded -1, 0, and 1, capturing the full range of responses. *Share UNIV among state employees* and *Share UNIV among peers* refer to canton-year averages of universalism endorsement among state and private sector employees, respectively. *Left-wing state* is a binary indicator equal to 1 if the cantonal executive is left-leaning (see text for coding details). Standard errors are clustered at the individual level. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

A.3 Mechanism for state normative effect

Table 10: Switchers to *vs.* stayers in the public non-HES sector

| Variable	(1) Switchers	(2) Stayers	(3) Difference
Share of female	0.429	0.351	0.077
	(0.497)	(0.478)	(0.054)
Share of married	0.623	0.676	-0.053
	(0.487)	(0.469)	(0.053)
Number of children	1.388	1.584	-0.197
	(1.289)	(1.310)	(0.141)
Age	38.714	40.512	-1.798*
	(9.467)	(8.666)	(1.021)
Share of Educ.: college level	0.143	0.242	-0.099**
	(0.352)	(0.428)	(0.040)
Share of living in urban area	0.612	0.574	0.038
	(0.490)	(0.495)	(0.053)
Ln yearly income	10.924	11.051	-0.127*
	(0.612)	(0.602)	(0.070)
State Universalism	0.679	0.729	-0.050
	(0.470)	(0.445)	(0.056)
Ideology	4.273	4.540	-0.268
	(2.263)	(2.107)	(0.258)

Notes: The table displays socio-demographic characteristics of switchers from the private into the public sector and stayers in the public sector within non-HES industries. The last column presents the difference in the mean value of each variable between the two sectors. Columns (1) and (2) standard deviation in parentheses. Columns (3) robust standard errors in parentheses. $*p < 0.10$, $**p < 0.05$, $***p < 0.01$.

Figure 10: Self-selection based on the universalism social norm across industries
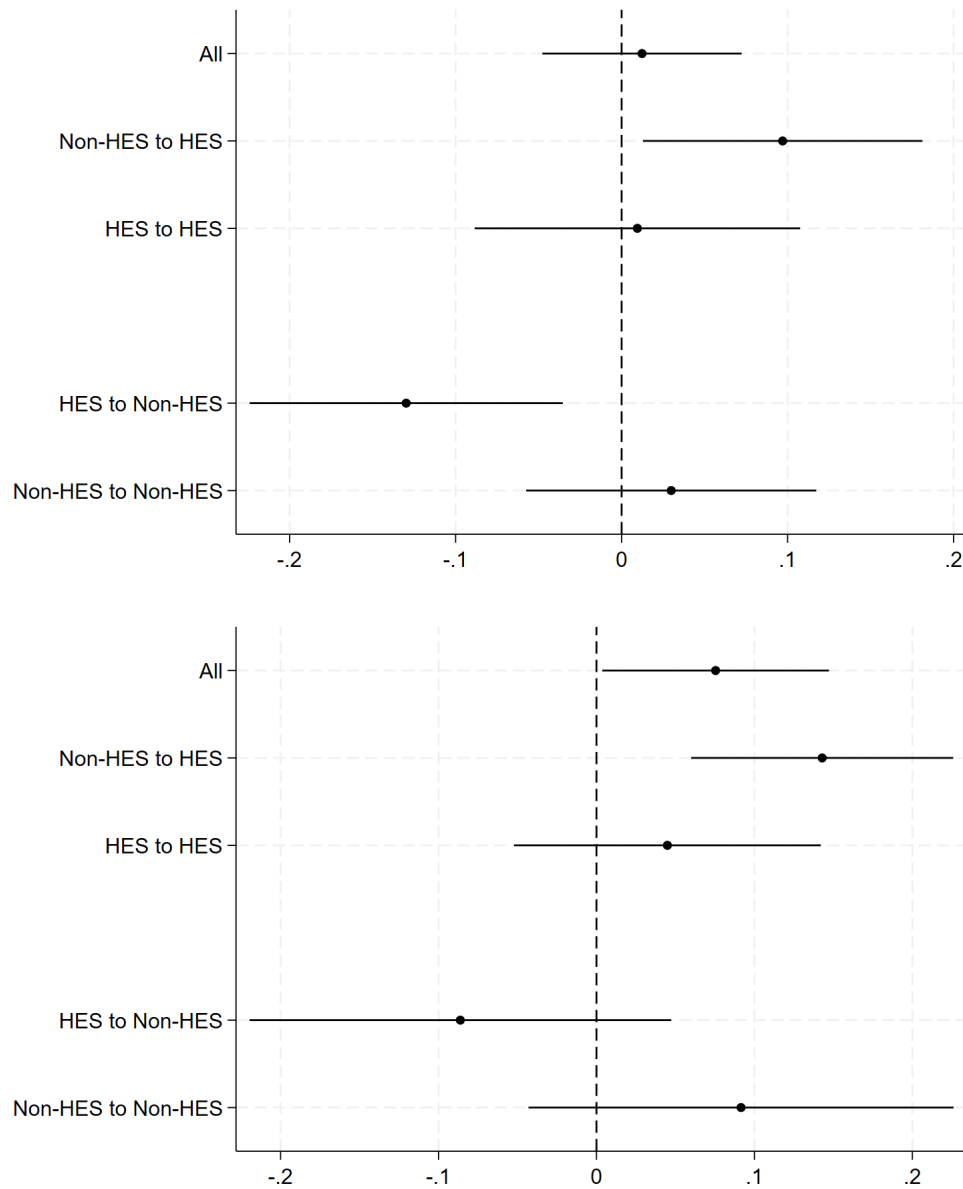
Notes: The figure summarizes regression coefficients of the switching indicators, capturing systematic differences in the universalism social norm between those who stay and those who switch at some future point. The above figure considers all switches while the figure below excludes switches at the federal level. See the text for further details. The 90% confidence intervals are based on standard errors clustered at the individual level.

Figure 11: Universalism social norm – robustness of working for the cantonal bureaucracy effect, dropping one cantonal bureaucracy at a time

Excluding cantonal bureaucracy one at a time
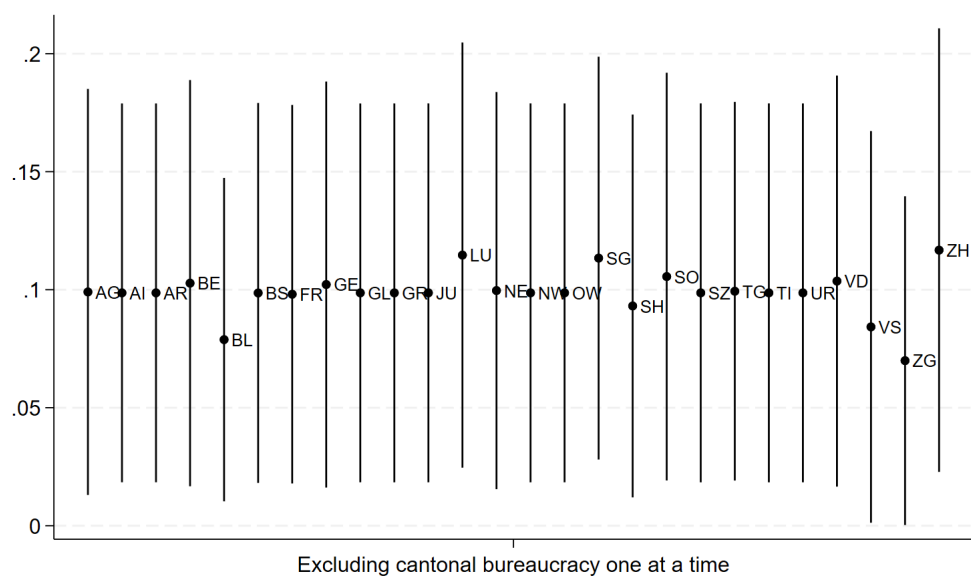
Notes: Figure reports point estimates of the state effect (specification from Column (4) in Table 5), considering switches within non-HES occupations, dropping one state at a time. Reporting 90% confidence intervals. Standard errors clustered at the individual level.