Alternative Worldviews, Distrust, and Populism*

Helios Herrera† and Federico Trombetta‡

March 28, 2024

Abstract

We define populism as a platform proposing a policy based on a mis-specified model of the world, a simple alternative worldview. Voters' trust in the mainstream political class evolves over time and depends on traditional politicians' performance once in office. Crucially, we assume this political distrust behaviourally increases the trust in simple alternative worldviews proposed by populists. Traditional politicians are aware of this risk and try to prevent it. Using this novel framework, we study when voters elect populist politicians, how populist and their alternative worldviews survive/recur and their long-run effect on several measures of voter's welfare. We show the link between distrust and alt-truth gives rise to "populist cycles" where trust never converges to its true level. However, cycles can be welfare improving for the voter. The increased pervasiveness of alt-worldviews is not always detrimental as it has a disciplining effect on mainstream politicians.

Keywords: Policy Complexity, Populist Trap, Behavioural Voters.
JEL Classification: D72, D83, D91

*We thank Federico Boffa, Massimo Morelli, Paolo Roberti, Stephane Wolton and participants to the Formal Theory Virtual Workshop, Unicatt Political Economy work in progress seminar, Bolzen Economics departmental seminar, EUI workshop "The Political Economy of Polarization and Policymaking", 2023 SIEP Conference and LISER workshop on "Migration and Politics" for their comments.

†Department of Economics, University of Warwick & CEPR

‡DISEIS, Università Cattolica del Sacro Cuore, Milano

1 Introduction

"The last two years of the pandemic have shown, in no uncertain terms, the importance of trust in science and politics, as well as in social discourse, but also the fragility of this trust. [...] [Our democracy] lives on solidarity and trust, and also trust in the facts."

Angela Merkel (2021)

1.1 Motivation

People's trust in politicians and political institutions can influence their attitudes toward experts and scientific information. For example, if people perceive politicians to be corrupt or dishonest, they may be more likely to view scientific information with skepticism or distrust, particularly if that information conflicts with their pre-existing beliefs or values. Figures 1 and 2 document a positive and strong correlation between trust in government and two different measures of trust in experts: the belief that science is good for society and the degree of confidence in universities.[1]

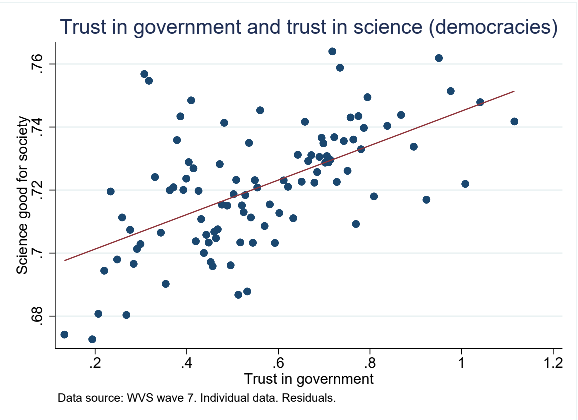

Figure 1: Trust in government and trust in science

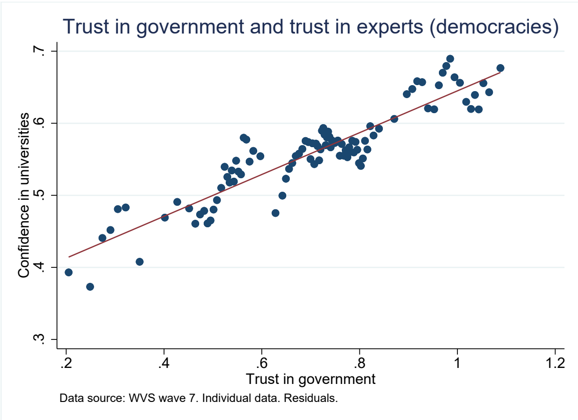

Figure 2: Trust in government and trust in experts

Distrust in experts, in turn, can facilitate the spread and promotion of alternative worldviews that are not supported by evidence. When people do not trust experts and

[1]Figures are based on WVS data, 7th wave. We use individual data, controlling for age, gender, income, employment, political position, country FE, survey mode FE and we plot binned residuals After controlling for the variables above, the correlation remains significant at 1 percent level. One standard deviation decrease in trust in government is associated with a 0.05 standard deviations decrease in beliefs that science is good for society, and with a 0.3 standard deviations decrease in confidence in universities.

scientific information, they may be more likely to believe in and promote alternative ideas, regardless of whether there is any evidence to support them. This can be particularly problematic when it comes to issues like public health or climate change, where scientific evidence is critical in informing policies and decisions. For example this graphs shows the correlation between distrust in science and in politics.[2]

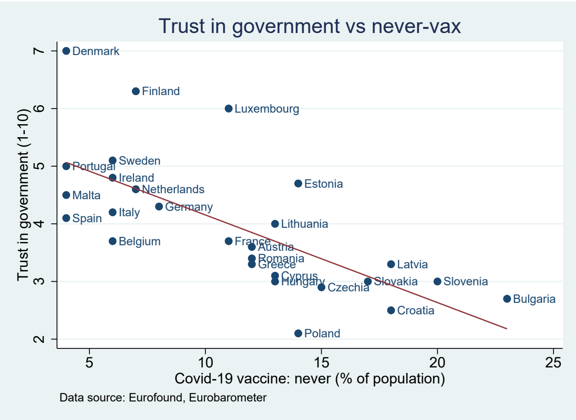

Figure 3: Trust in government and anti-vax sentiment

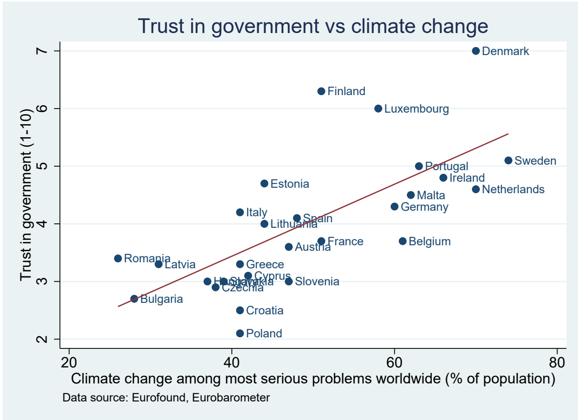

Figure 4: Trust in government and importance of climate change

Crucially, alternative worldviews can also have an impact on economic policy and lead to the promotion of counterproductive policies that may have negative consequences in the long run. When people reject mainstream economic theories and the advice of experts, they may be more likely to support policies that are based on alternative ideas or ideologies, even if those policies are not supported by evidence or may have unintended consequences.

Thus, when mainstream or traditional politicians are perceived as corrupt or out of touch with the public, it can create a political opening for opportunistic politicians who promote alternative worldviews that may be at odds with sound economic policy.

1.2 Populism and Alternative Worldviews

We define populist a party running on simple alternative worldviews. We use the word populism as we think this feature captures important elements of what populist politicians

[2]We use Eurostat, country level data.

do in the real world.[3] Populist politicians often portray themselves as champions of the people offering simple solutions to the problems that mainstream politicians have failed to address (Guriev and Papaioannou, 2022). They often advocate for extreme policies not supported by evidence but based on ideological or emotional appeals rather than sound economic principles, and receive electoral support for those proposals. Figures 5 and 6 show that there is a negative correlation between trust in science or in experts and the degree of populism of the party voted by individuals.[4]

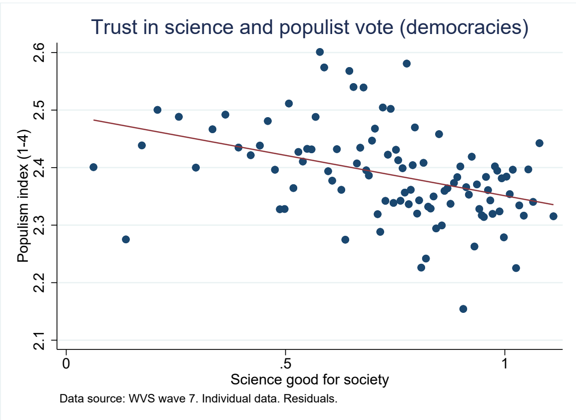

Figure 5: Populist vote and trust in science

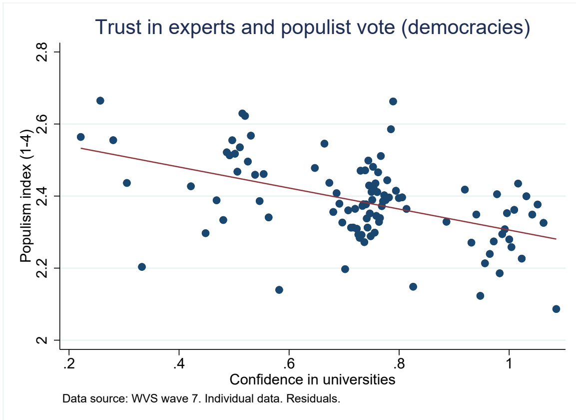

Figure 6: Populist vote and trust in experts

However, while populist politicians may initially gain popularity by appealing to public frustration and dissatisfaction with the status quo (Silva and Wratil, 2023), their simple drastic solutions may ultimately have negative consequences for the economy and society as a whole. For example, trade and environmental policy issues are complex and often do not have straightforward policy solutions. These issues involve a delicate balance between promoting economic growth through international trade and protecting domestic jobs as well as the environment. Simple populist policies that promote protectionism or isolationism may be popular among some groups, but may lead to vast economic inefficiencies and reduced growth in the long run. Similarly, policies that ignore environmental externalities or that prioritize short-term gains over long-term sustainability may lead to

[3]Obviously, there are other important aspects of populism that we do not capture here. Different definitions of populism have been used in the literature.

[4]We use the same individual data and specification from the WVS as in figures 1 and 2. Again, the correlation is strongly significant and it implies that a one standard deviation decrease in the belief that science is good for society (in confidence in university) increases the degree of populism of the voted party by 0.03 (0.05) standard deviations.

negative consequences for both the economy and the environment.

This phenomenon has become much more prominent in recent years. In the past, traditional media outlets served as gatekeepers of information, largely controlling the narrative and deciding what stories were worth covering. Now the democratization of information and the rise of social media has enabled individuals to access and disseminate information more easily than ever before. Populist politicians often use social media to communicate directly with their followers, bypassing traditional media outlets and appealing to emotions and grievances rather than reasoned arguments. In fact, broadband diffusion boosted support for populist parties (Campante et al., 2018; Guriev et al., 2021). This has allowed them to build large followings and mobilize supporters more effectively than traditional political parties, spreading alternative worldviews, known to circulate more on social media (Vosoughi et al., 2018).

1.3 Our Take

Economic policies such as immigration, trade, climate policy, are complex and not easy to address: there is a fractured policy consensus on globalisation, border openness and climate change which cuts across the left-right divide. Indeed, though culture cleavages remain, left and right have largely disappeared as far as economic policy goes. But people are rightly worried that politicians may use the complicated worldview as a smokescreen to hide their ulterior motives and manipulate policies to their benefit. Indeed, when the issues are complex and constantly require a balancing act the inference is imperfect. This, in turn, generates legitimate distrust on mainstream politicians.

Our goal is to build a model to try to shed light on the interaction between traditional policy, which is a complex constant compromise and may often be biased, and the surge persistence and recurrence of populism which offers more extreme policies based not based on evidence but on simple alternative worldviews. We aim to shed light not only on the surge on populism, but most importantly on what happens after populism, and how this phenomenon may recur, analysing its long term welfare consequences.

We build a dynamic, infinite horizon model, where a voter selects the politician in power in every period, drawing from a fixed pool. Traditional politicians may be biased, i.e. willing to pursue policies misaligned with the interest of the voters. We model trust in traditional politics as voters' updated beliefs on the share of unbiased politicians in the pool. Our novel behavioral assumption is that lower trust implies a higher probability that the voter buys the "populist worldview", i.e. a mis-specified model of the economy proposed by the populist platform. In particular, the populist worldview is the simplest explanation of reality available, that excludes all the nuances of policymaking. As everything is simple, in this worldview there is only one correct policy choice, and politicians cannot exploit any uncertainty at their advantage. Traditional politicians are aware of this populist threat and try to prevent it. In our definition, populism is the political strategy of proposing an extreme alternative worldview and a policy consistent with it.[5]

1.4 Results

We show that populism can be an equilibrium strategy, namely it is optimal from the point of view of a rational candidate willing to maximize her party's electoral prospects. Moreover, in equilibrium, when trust is moderate and extreme alternative worldviews are sufficiently effective, populists can discipline traditional politicians, reducing their incentives to choose policies misaligned with voters. While this is beneficial for welfare, it partly confounds voters' learning on the nature of traditional politicians and implies a positive chance of falling into populism.

In some cases, voters' trust may fall below a threshold generating a "populist cycle", in which traditional and populist politicians keep alternating in office and trust in the traditional political class is never restored. This populist cycle is a trap, namely an absorbing state for the trust state variable. Traditional politicians cannot regain the people's trust they lost, because the people cannot tell if politicians are trustworthy and aligned with them or biased and acting to defeat the populist threat.

We conduct a welfare analysis assuming the optimistic view that (some) traditional

[5]This idea of populists proposing simple worldviews is consistent with several recent papers, such as Levy et al. (2022), Carillo et al. (2023) and Andreottola and Sartori (2023).

politicians deserve the voter's trust. In this case we have two possible steady states, one in which there is convergence to the truth and traditional politicians stay in power as often as possible, and another in which traditional and populist politicians alternate in power and trust does not converge to the truth. Perhaps surprisingly, in terms of steady state welfare this populist cycle may be better than the steady state convergence to the truth. This is due to the trade-off triggered by the presence of populist politicians: they may induce more discipline from traditional politicians, but this works only as long as the election of (inefficient) populists is a sufficiently likely possibility. To paraphrase a popular saying, voters jump out of the traditional frying pan into the populist fire. Yet, the presence of the fire may tame the heat of the frying pan.

The new media environment has indeed played a significant role in facilitating the spread of alternative worldviews and the surge of populism around the world. A key comparative statics variable is the extent to which distrust in traditional politics makes voters more gullible to alternative worldviews. We show that sufficiently pervasive alternative worldviews (i.e. a sufficiently high α in the model) makes the populist a credible alternative to (possibly biased) traditional politicians. For intermediate levels of trust, this induces moderation from biased traditional politicians (i.e. they choose the wrong policy less often). An increase in α can have a non-monotonic effect on the probability of transitioning from a traditional to populist politicians. As mentioned above, in terms of true steady state welfare of the voter, an increase in α need not be bad: the moderating effect may more than compensate the increased probability of electing a populist. We shed light on when and why welfare in the populist-traditional cycle is better than welfare in the steady state where trust in politicians is high. This happens when disciplining those that are biased is particularly important, namely when the share of "good" traditional politicians in the good state is sufficiently low, as might be the case in some countries. In a key extension to the model, we endogenize the entry of a populist challenger. We show that, if there is a positive fixed cost to entry for non-traditional parties, then with higher α Populist will enter sooner and we explore how this changes welfare. The main results of the paper are qualitatively unchanged: populist entry happens for sufficiently high α and this may lead to a cycle where trust never converges to its true value. Those cycles can be welfare improving for the voter. Importantly, we can also show that, when the steady state true welfare of the voter is higher in the cycle than when trust converges to the truth, then an increase in α can be good for the ex ante long term welfare as well.

1.5 Related Literature

Our paper contributes to the growing theoretical literature on the political economy of populism and its determinants (Prato and Wolton, 2018; Crutzen et al., 2020; Morelli and Sasso, 2021; Levy et al., 2022; Bernhardt et al., 2022; Gratton and Lee, 2023a; Carillo et al., 2023). As in Levy et al. (2022), the driving force is the adoption of misspecified models of the world. However, the type of misspecification we look at is different, and we consider our analysis in a dynamic accountability model. We share with Bernhardt et al. (2022) the importance of the dynamic structure of the model and the possibility of cycles. However, the type of cycle we find is different, and we consider a model where learning is the driving force. We share with Carillo et al. (2023) the assumption that populist politicians propose "simpler" policies than their traditional counterpart, and we motivate this idea with their proposal of a simpler worldview. We share with Morelli et al. (2021) the accountability structure and the idea of populists being committed to a certain policy, and we add an endogenous trust-driven dynamics and worldviews acquisition. Crutzen et al. (2020) and Bordignon and Colussi (2020) also consider endogenous populism, but they do so in a static model of electoral competition. Furthermore, differently from all of them (except Bernhardt et al. (2022)), we study the "what's next" part, trying to understand what follows populist entries and, potentially, populist victories. In this sense, we take a similar dynamic approach as in Gratton and Lee (2023b), whose focus is on rise and fall of "illiberal democracies", rather than on the presence (or absence) of populist politicians in power. The idea of populist politicians as a way to discipline traditional politicians is also in Auriol et al. (2023), where voters are fully rational. Our setting, however, is quite different, as the driving force is based on a combination of adverse selection and moral hazard (rather than just moral hazard) motivated by the voter's updating on the level of trust (which, in our model, has specifically an adverse selection component. In Auriol et al. (2023), instead, it is about the motivation behind a certain action, but there is no asymmetric information on politicians' types). As a consequence, the crucial state variable is different and we show the possibility of cycles. Finally, Svolik (2013) highlights that bad outcomes from the government lead voters to update negatively on politicians, thus making it easier for "bad" politicians to join the pool in a self-fulfilling trap of democratic breakdown. We obtain a similar starting point under different conditions (there is no costly monitoring from voters and the pool of politicians does not change). But then our "trust-driven" dynamics, and the way in which populist politicians may be an effective disciple device, suggests instead that the polity may end up into a cycle, where beliefs updating remains steady in the long run.

On the technical side, the way we model worldviews is similar to Ash et al. (2021). Differently from them, our paper considers endogenous populism and a dynamic environment. As Little (2019) and Bowen et al. (2023), we consider a model where learning happens in a (possibly) misspecified setting, although in a different setting. We assume that voters have limited memory, as in Levy and Razin (2021), and that players care only about the period ahead as in Forand (2021).

More broadly, the results of our dynamic model of accountability (Duggan and Martinelli, 2017) are consistent with a series of other papers on the effects of "rationality" or "information" on voters' welfare and behaviour in political agency, showing that sometimes less information / less "rationality" can be good (Blumenthal, 2022; Prato and Wolton, 2018; Trombetta, 2020; Ashworth and Bueno De Mesquita, 2014).

On the empirical side, several papers are studying the economic and cultural determinants of populism (Guiso et al., 2020, 2019; Margalit, 2019; Guriev and Papaioannou, 2022; Boffa et al., 2023). Our theoretical results help explaining some recent empirical results showing that bad performance during the recent Covid crises increases voters' distrust in the political class (Martinez-Bravo and Sanz, 2023; Becher et al., 2021). Consistently with our model, Boffa et al. (2023) shows that the bad performance of incompetent local politicians increases the vote share of populist parties in national elections.

Finally, our assumption in traditional politicians being informed but potentially biased is consistent with Bombardini et al. (2023) findings on American politicians' information about the China shock.

2 Model

We consider an infinite horizon game, with periods denoted by $t = 0, 1, 2....$ There are three players in every period: two politicians and a representative voter (ν). All of them are replaced in every period. In every period, a policy relevant state of the world $\theta_t \in \{0, 1\}$ is drawn from from the same distribution. The true probability that the state is 0 assumed to be $\lambda^T \in (0.5, 1)$. For example, θ_t can be interpreted as whether social welfare is maximised by lower (0) or higher (1) barriers to trade, or more generally whether globalization is good or bad, and so on. Note that in this framework λ^T captures the uncertainty over the policy relevant state, or more generally the complexity versus the "one-sidedness" of the policy choice.

In every period, two politicians, one populist and one traditional, face each other in an election, that determines who is in power for that period. In Appendix B we consider an extension of the model with two traditional politicians, showing the existence of equilibria with similar features.[6] All politicians are drawn from the same pool, whose characteristics are fixed throughout the game. In particular, every politician i has a true type $\Gamma^i \in \{G, B\}$ and the true share of politicians of type G in the pool is dented by γ. We assume that $\gamma \in \{0, z\}$: the set of realisations is known to every player, but the true γ is not. We denote by τ_t players' beliefs on the probability that the pool of politicians contain a share z of G types, updated with the information set reached at that point in the game (Ω_t). Formally, $\tau_t = Pr(\gamma = z|\Omega_t)$. Intuitively, politicians can be individually good (G) or biased (B). Overall, the true share of good politicians is γ and the evolution of players' beliefs abut whether there are good politicians in the pool is captured by τ_t, a crucial element of trust in politicians. Overall, we model trust as $z\tau_t$, i.e. the expected share of G politicians in the pool measured with the information known at time t.

[6] In Appendix C we endogenize the entry of the populist politician, showing that all our results hold.

Once in power, the politician who wins the election learns θ_t and chooses the policy $x_t \in \{0, 1\}$. Votes and good politicians (when in office) derive a policy-payoff of 1 if $x_t = \theta_t$ and 0 otherwise. We define u_t the observed outcome of the policy choice, with $u_t = 1$ if $x_t = \theta_t$ and 0 otherwise. Biased politicians derive a policy-payoff of 1 (when in office) if and only if $x_t = 0$. Intuitively, we assume that biased politicians are biased in favour of openness to trade. Furthermore, politicians derive an additional payoff of E if the party they belongs to wins the election in the next period.

2.1 Elections and worldview

Elections are at the beginning of every period and they are influenced by the worldviews proposed by the existing candidates. We model worldviews as different (and possibly wrong) probability distributions over θ_t, defined by λ_t. In particular, we call "traditionals" those politicians that adopt a true worldview, stating that $\lambda_t = \lambda^T$. On the other hand, we call "populists" those politicians proposing a simple/extreme alternative model of reality, one where $\lambda_t = 0$ (e.g. trade openness or climate protection is always bad). Note that claiming that $\lambda_t = 0$ implies ignoring all the complicated nuances of reality. The true worldview says a certain policy works sometimes, depending on the circumstances. The "populist" worldview claims that those complications are far-fetched and that reality is simple, as the probability distribution over the state space is degenerate. As a consequence $x_t = 1$ is the obvious optimal policy. In a nutshell, alternative worldviews are misspecified models of reality. We use the world "identity" and the letter $s \in \{T, P\}$ to refer to whether politicians are traditionals of populists.

We make two crucial behavioural assumptions about voters. First, they are receptive to the alternative worldview in a way that is inversely proportional to trust in politics. More formally, we assume that, when politicians propose different worldviews, voters believe in the "populist" one with probability proportional to voter's distrust for traditional politics, namely: $\alpha(1 - z\tau_t)$. $\alpha \in [0, 1]$ is a parameter that captures the importance of our behavioural assumption, i.e. how strongly distrust in T politicians translates into accepting a different worldview. Intuitively, we are assuming that low trust in politicians (i.e. the belief that many of them are biased in favour of the "elite") opens the avenue for populists offering alternative worldviews, that can be believed with some probability. Second, we assume that voters have limited memory, meaning that they condition their choices on the state variable only (i.e. on τ_t and u_{t-1} when relevant), without trying to infer, from the timing, information on whether populists have been in power in the past, were effective and so on.

2.2 Timing

At the beginning of the game, Nature chooses the true quality of the pool of politicians and τ_0. Then, in every period:

1. The state variable τ_t is determined (equal to the posterior from the previous period);
2. ν chooses the worldview λ_t and votes;
3. The winning politician learns Γ^i, θ and chooses x_t (constrained to 1 if $s = P$);
4. u_t is observed and ν calculates the posterior $\tau_{t+1}(u_t)$;
5. Period ends;

All players are myopic, but we assume that party members care about the electoral perspectives of their party in the next period (through E).

The solution concept is Perfect Bayesian Nash Equilibrium. We focus on equilibria where producing the good outcome is not punished reputationally.

2.3 Discussion

We assume that the adoption of a different worldview does not imply a re-evaluation of the past information available. Voters adopt the worldview and calculate their expected payoffs from the alternatives they have based on that. Second, we assume that voters revert back to the correct worldview once the populist fails to deliver (conditional on the populist worldview, that is a zero probability event), but they may be fooled again in the future because of the limited memory. This is broadly consistent with the idea of selective memory triggered by prominence, as outlined by Bordalo et al. (2020): populist failures fade away after the first period and they are no longer used in the decision making. Finally, in this analysis we focus mostly on parameters such that the true γ is equal to z and τ_0 is high. Thus, an optimistic scenario.

3 Within-periods Analysis

The last choice happening in every period is the policy choice by the incumbent. It depends on the anticipation of the results of the elections in $t+1$, hence its analysis is a natural starting point. We divide it into different sub-cases, showing what happens to ν's electoral choice (in $t+1$) and to the policy choice of the politician in power in period t depending on the identity of the politician in power.

3.1 Traditional in power

V's choice is based in two elements: the "inherited" τ_{t+1}, i.e. the beliefs over the quality of the pool of politicians, and the worldview she chooses to adopt. Generally speaking, ν derives an expected utility from electing the T candidate defined as

$$\mathbb{E}(u_{t+1}(T)|\tau_{t+1},\lambda_{t+1}) = \lambda_{t+1} + (1-\lambda_{t+1})\left[z\tau_{t+1} + (1-z)(1-\sigma^c_{t+1})\right] \tag{1}$$

where σ^c_{t+1} denotes the conjectured probability of policy distortion (that will be correct in equilibrium) by a B politician. Similarly, the expected utility from electing the P candidate is

$$\mathbb{E}(u_{t+1}(P)|\tau_{t+1},\lambda_{t+1}) = 1-\lambda_{t+1} \tag{2}$$

It is straightforward to note that (2) is weakly below λ_{t+1}. Hence, when the adopted worldview is λ^T, ν chooses the T candidate. *Viceversa*, when the adopted worldview is 0, ν chooses the P candidate (If $\sigma^c_{t+1} = 0$ and $\lambda_{t+1} = 0$, the voter is actually indifferent between the P and the T politician. We assume that this indifference is broken in favour of P). The next Lemma summarizes the election choice, proofs are in Appendix A.

Lemma 1 *In every period $t+1$ where one candidate runs with identity T and the other with identity P, we have that ν chooses the T candidate if $\lambda_{t+1} = \lambda^T$ and the P candidate otherwise. The ex-ante probability of a victory of the P candidate is $\alpha(1-\tau_{t+1}z)$.*

Given this result, we can now move back to the policy choices in period t. G politicians always choose the policy that matches the state, and the same is true for B politicians when $\theta_t = 0$. This is due to the fact that, for any x_t, the chances of the incumbent's party to win in $t+1$ against a P challenger are increasing in τ_{t+1}, which is (weakly) higher when the policy produces good outcomes. Hence, only B politicians when $\theta_t = 1$ face a trade-off, but its solution is complex. First, note that, as a T is in power, ν holds correct beliefs over λ, at least in period t. Hence, for any conjectured σ_t^c, ν's updating after observing u_t will be as follows:

$$\tau_{t+1}(u_t = 1) = \frac{\tau_t(1-(1-z)(1-\lambda_t)\sigma_t^c)}{1-(1-\lambda_t)\sigma_t^c(1-\tau_t z)}$$

$$\tau_{t+1}(u_t = 0) = \frac{\tau_t(1-z)}{1-\tau_t z}$$

where the second equation is defined as long as $\sigma_t^c \neq 0$. Note however that, sometimes, $u_t = 0$ is off path (or conjectured to be off path). We assume that the updating works in the same way as in the on-path case. This is equivalent to assuming that, when $u_t = 0$ is off path, if ν observes $u_t = 0$ she assumes that it comes from a biased individual politician, hence updates negatively on τ.

It is useful to briefly consider the different effects that λ^T and z have on the posteriors, summarized in Lemma 2.

Lemma 2 *For any fixed $\sigma_t^c > 0$, $\tau_{t+1}(u_t = 1)$ is increasing in z and decreasing in λ^T. $\tau_{t+1}(u_t = 0)$ is decreasing in z and does not change with λ^T.*

Intuitively, z increases the distance between the two realizations of the posterior because it increases $Pr(u_t = 1|\gamma = z)$ without affecting $Pr(u_t = 1|\gamma = 0)$. Hence, it makes the policy outcome a better signal of the overall quality of the pool. The effect of λ^T is the opposite: by increasing the probability that the bad politician is aligned with the voter, it decreases the ability of $u_t = 1$ to carry information about the pool: that outcome is increasingly likely to come from the bad pool as well.

Because of Lemma 1, the T incumbent knows that the probability that his party remains in power in the next period is $r_{t+1}(u_t) = 1 - \alpha(1 - z\tau_{t+1}(u_t))$. Hence, when $\theta_t = 1$, a B politician faces a trade off between choosing $x_t = 1$, that guarantees a payoff of $r_{t+1}(u_t = 1)E$ or $x_t = 0$, that gives a payoff of $1 + r_{t+1}(u_t = 0)E$. Furthermore, in equilibrium σ_t^c must be correct. Those elements allow us to pin down the equilibrium:

Proposition 1 *Define* $\Delta\tau_{t+1}(\sigma_t) = \tau_{t+1}(u_t = 1, \sigma_t) - \tau_{t+1}(u_t = 0, \sigma_t)$. *There exists a PBNE where* σ_t *is as follows:*

1. $\sigma_t = 1$ *if* $1 \geq \Delta\tau_{t+1}(\sigma_t = 1)\alpha z E$*;*
2. $\sigma_t = 0$ *if* $1 \leq \Delta\tau_{t+1}(\sigma_t = 0)\alpha z E$*;*
3. $\sigma_t = \frac{1-\tau z(1+(1-\tau)\alpha z E)}{(1-\lambda)(1-\tau z)^2}$ *otherwise;*

Intuitively, in equilibrium the level of distortion has the shape described by figure 7: there is full distortion for low τ_t and high τ_t, where the outcome does not affect too much the election chances, and possibly no distortion and mixed equilibria in between. In order to guarantee the existence of a range of τ_t where in $\sigma_t = 0$, we need E to be sufficiently high. Otherwise, it may be that $\sigma_t = 1$ always or that it is either 1 or the mixed strategy described below. The condition for having a range of τ where $\sigma = 0$ is an equilibrium is $E > \frac{2-z+2\sqrt{1-z}}{\alpha z^2}$.

3.2 Populist in power

Finally, suppose the winner of period t elections was running with a P identity. In this case, there is no updating on the pool following the policy choice and the outcome, because populists are committed to $x_t = 1$. As ν believes $\lambda_t = 0$ and there is a commitment to choose $x_t = 1$, there cannot be any $u_t = 0$ outcome, if that one was the correct model

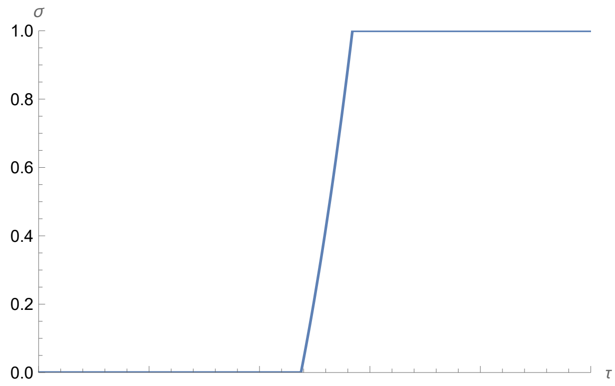

Figure 7: Equilibrium probability of distortion, conditional on a P challenger, as a function of τ_t for sufficiently big E. In particular, $E = 8$, $\alpha = 1$, $z = 0.6$, $\lambda^T = 0.85$.

of the world. We assume that, in this case, ν's beliefs revert back to the true worldview, hence to $\lambda_t = \lambda^T$. This has a direct consequence on the survival rate of a populist in power: a populist incumbent loses office with probability λ^T. Under those assumptions, the evolution process of λ_t depends also on the identity of the incumbent.[7]

The probability of a populist losing power is distributed as a Geometric random variable. As a consequence, we can calculate the expected duration of a populist government.

Observation 1 *The expected duration of a populist government is* $\frac{1}{\lambda^T}$.

Furthermore, note that this model is characterised by a form of "populist disappointment": the expected performance of populists in office, from the voter's side (hence conditional on believing to the worldview $\lambda_t = 0$) is better than the actual one.

Observation 2 *When P is in office, the voter over estimates the probability of receiving a good outcome, vis-a-vis the true probability.*

Observation 2 originates from the fact that $Pr(u_t = 1|P, \lambda_t = 0) > Pr(u_t = 1|P, \lambda_t = \lambda^T)$. Intuitively, ν chooses a P politician when she believes in the populist worldview. But this implies that she is systematically overestimating the probability that a P politician delivers the good outcome. Hence, in a sense, populists can "surprise" the voter in a negative way.

[7] As shown in Appendix D, we get very similar results even if the choice about λ_t follows the same probability distribution described above irrespective of whether there is a T or a P incumbent.

4 Dynamics

In this section we study the dynamic evolution of the model, and the evolution of trust in equilibrium. We first consider the benchmark case of $\alpha = 0$, i.e. where our behavioural assumption is not in place. Then, we compare it with the case of a sufficiently high α, showing that in this case the populist cycle can appear in equilibrium. Third, we present few simulations of the model, to get a better understanding of the comparison between the benchmark case and the case where α plays a role. Finally, we derive some comparative statics on the transition probabilities, i.e. on the per-period probability of switching from a traditional to a populist government and *viceversa*.

4.1 Benchmark: $\alpha = 0$

If $\alpha = 0$, P is never chosen and the bad T politician chooses $\sigma = 1$. Over time, the posterior of the voter (defined as $\hat{\tau}$) will converge to the true state: conditional on the "good pool", $\hat{\tau}$ converges to 1, and conditional on the "bad pool" $\hat{\tau}$ converges to 0. To see this, note that from the point of view of the voter, that correctly anticipates $\sigma = 1$, we can write down the posterior, for a generic prior τ_0, after observing n realisations of the outcome, k of which are =1 (and the remaining =0).

$$\hat{\tau}(k,n) = \left(1 + \frac{1-\tau_0}{\tau_0}\left(\frac{\lambda^T(1-z)}{z(1-\lambda^T)+\lambda^T}\right)^k \left(\frac{1}{1-z}\right)^n\right)^{-1} \tag{3}$$

It is clear from above that the sequence of good and bad outcomes does not matter for the definition of the posterior. For any given value of the parameters, the posterior depends on the number of positive or negative outcomes, irrespective of the sequence in which they come. By the law of large numbers, this converges to the true state.

We can also determine a sufficient condition on α such that, for α below this level, the sole equilibrium of the game is one where $\sigma = 1$.

Proposition 2 *If $\alpha = 0$ or sufficiently small, in the unique equilibrium of the game there is full distortion.*

Proposition 2 shows that there is a threshold on α, bounded away from zero, such that full distortion (i.e. $\sigma = 1$) is the only equilibrium of the game, for any combination of the other parameters. We define this threshold $\bar{\alpha} := \frac{4\lambda^T(1-z)}{zE}$. Intuitively, if the contagion between distrust and alternative worldview adoption is weak, biased traditional politicians do not have incentives to moderate their behaviour.

Over time, because of Martinagle's convergence theorem, learning converges to the true state of the pool of politicians, slowing down just in the relatively rare periods with a P in office.

4.2 Dynamics with high α

In this section, we assume parametric restrictions to focus on parameters range where results are interesting and the model is tractable. Hence, we assume that α and E are sufficiently big so that there is a nonempty set of τ_t such that $\sigma_t = 0$. Let us define the extremes of this interval as $\underline{\tau}^0, \bar{\tau}^0$, noticing that we can find a closed form for those boundaries. In particular, $\underline{\tau}^0 = \frac{1+z\alpha E-\sqrt{(1+z\alpha E)^2-4\alpha E}}{2z\alpha E}$ and $\bar{\tau}^0 = \frac{1+z\alpha E+\sqrt{(1+z\alpha E)^2-4\alpha E}}{2z\alpha E}$. One important point to notice is that $\tau_{t+1}(u_t = 1) = \tau_t$ when $\sigma_t = 0$, meaning that there cannot be updating on trust without some positive probability of distortion. We are now in the position to assess the dynamic evolution of the system.

Suppose $\tau_0 > \bar{\tau}^0$ and sufficiently high so that that $\sigma_0 = 1$. If the incumbent politician is of B type and $\theta_0 = 1$, the outcome will be bad, i.e. $u_0 = 0$. As a consequence, $\tau_1 < \tau_0$. Obviously, if $\theta_1 = 0$ then trust move upward, and so on. Suppose however that τ_1 is still sufficiently high so that $\sigma_1 = 1$ and, if $\theta_1 = 0$ and the incumbent politician is of B type, $u_1 = 0$. Hence, $\tau_2 < \tau_1$. If this process continues, there is a positive probability to reach $\tau_t < \bar{\tau}^0$. As soon as this happens, we have $\sigma_t = 0$. As a consequence, $u_t = 1$ for every θ_t (if T is in power), but $\tau_{t+1} = \tau_t$. This, combined with the fact that τ_t does not move when the incumbent is a populist, implies that this condition is absorbing: irrespective on who is in power, $\tau_{t+1} = \tau_t$; T incumbents behave well, even if biased, but trust is not restored; the incumbent switches from T to P with probability τ_t and from P to T with probability λ^T. A P challenger remains such even after losing, as he may gain power

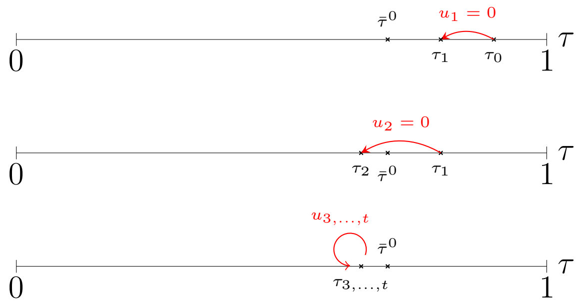

Figure 8: One possible path for the evolution of τ from τ_0 to the cycle. When τ_t moves below $\bar{\tau}^0$, i.e. after two consecutive negative outcomes, power keeps cycling between T and P and trust remains fixed at τ_2 forever.

again in the next period (with probability τ_t). We call this situation a "cycling" regime, whose path of entry is described in Figure 8.

Note that a direct consequence of the cycle is the fact that the system does not converge to the true state. Once populists kick in, trust cannot be restored endogenously. Proposition 3 summarizes this dynamics.

Proposition 3 *There exists a range in τ such that, if τ_t enters in this range, then τ remains constant for every following period, irrespective of the identity of the government.*

4.3 Simulations

To get some insights on this dynamic, we simulate the behaviour of τ over time as induced by the strategies of T politicians. When the P politician is in power, τ remains flat. Hence, a complete simulation of τ should include also flat periods in the graphs with α positive. All the graphs that follow, we set $\tau_0 = 0.9$, $E = 5$, $z = 0.8$ and $\lambda^T = 0.6$. We run 200 simulations each 80 periods long, conditional to the good state of the political pool (i.e. $\gamma = z$). The bad state works in a very similar way.

Figure 9 shows that, over time, trust moves toward the truth (i.e. $\tau = 1$) when $\alpha = 0$.

If we allow for a sufficiently high α so that P becomes a serious challenger for some values of τ, we see two different scenarios. On the one hand, trust may move toward the truth. However, still we may enter in the "cycling" regime with positive probability. τ tends to move upward, but there may be enough negative realizations that push it below

$\bar{\tau}^0$. Then, it stays there, as shown in figure 10.

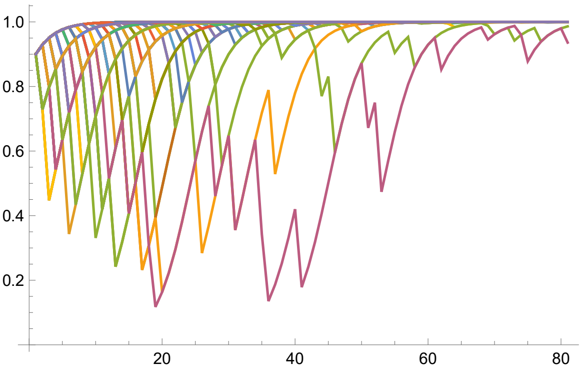

Figure 9: Evolution of τ as induced by T over time when $\alpha = 0$.

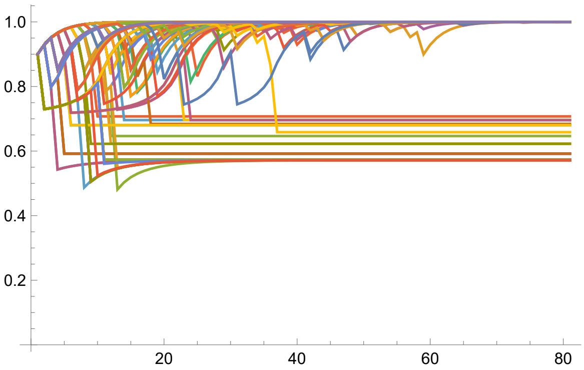

Figure 10: Evolution of τ as induced by T over time, $\alpha = 1$.

4.4 Transition probabilities

We are now in the position to study how transition probabilities are affected by the parameters of the model. We define them as follows:

$$Pr(PT) = \lambda^T \tag{4}$$

$$Pr(TP) = \alpha(1 - \tau_{t+1}(u_t = 1)z) + \alpha(1 - \gamma)(1 - \lambda^T)\sigma_t z \Delta\tau \tag{5}$$

Equation (4) is time invariant and depends just on the probability that the populist policy fails. Equation (5) depends on the true quality of the pool (i.e., the value taken by γ, either z or 0). Proposition 4 summarizes the main comparative statics results.

Proposition 4 *The effect of α on $Pr(TP)$ is as follows:*

- *If parameters are such that $\sigma \in \{0, 1\}$, then $\frac{\partial Pr(TP)}{\partial \alpha} > 0$.*
- *If parameters are such that σ is interior, there exists a region of parameters where $\frac{\partial Pr(TP)}{\partial \alpha} < 0$. Otherwise, $\frac{\partial Pr(TP)}{\partial \alpha} \geq 0$*

The effect of λ^T is as follows:

- *$Pr(PT)$ is increasing in λ^T;*
- *$Pr(TP)$ is weakly decreasing in λ^T if $\gamma = 0$ and weakly increasing if $\gamma = z$.*

Inspecting (5), it is clear that the role of α is not obvious. When parameters are such that σ is a corner solution, α unambiguously increases the chances of a transition from a T incumbent to a P incumbent. But things are different when σ is interior: α decreases σ, as shown in Proposition 1, and this as a consequence decreases both $\Delta\tau$ and $\tau_{t+1}(u_t = 1)$. Intuitively, α acts on the transition probability trough two channels. A direct one, increasing the chances that the populist worldview is accepted by the voter, and an indirect one, increasing the discipline of the T politician. This indirect channel, however, is active only if σ is a function of α, i.e. when it is interior.

Proposition 4 is also suggesting of a non-monotonicity in the effect of α on $Pr(TP)$, as illustrated in Figure 11. When α is low, $\sigma = 1$ (for that particular level of τ), hence $Pr(TP)$ is increasing. However, for intermediate α, σ is interior and the negative, indirect effect of more discipline dominates the direct effect of higher populist electability, decreasing the overall probability of transition. As α increases even further, however, the equilibrium σ becomes 0 and the effect of α is once again positive.

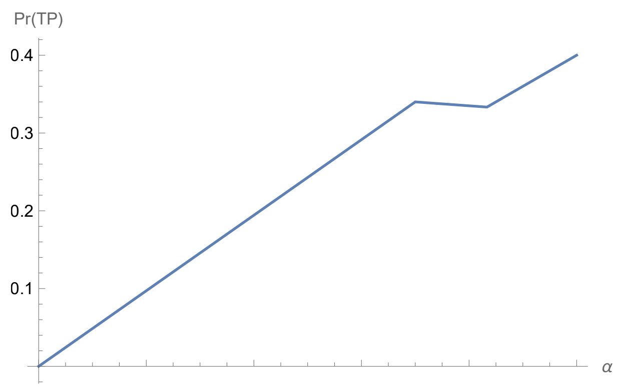

Figure 11: Effect of α on $Pr(TP)$ for $\tau_t = 0.75$, $z = 0.8$, $\lambda^T = 0.6$, $E = 4$, $\gamma = 0$.

Moving to λ^T, its effect on $Pr(PT)$ is obvious: the more likely it is the aligned state, the more likely it is that the populist will deliver a bad outcome and hence will be replaced. The effect on $Pr(TP)$ is more complicated. First, when $\sigma = 0$ and when σ is interior the effect of λ^T disappears. Second, inspecting equation (5) and recalling Lemma 2, it is clear that, when $\sigma = 1$, λ^T has several competing effects. On the one hand, a direct negative effect. Furthermore, it also decreases $\Delta\tau$. But there is a third (indirect) effect through $\tau_{t+1}(u_t = 1)$: it is decreased by λ^T, and this increases $Pr(TP)$. Note that both the negative effects are mediated by $1-\gamma$. Thus, when γ is small, they dominate. When $\gamma = z$, however, the positive effect dominates and an increase in λ^T increases the probability of switching from T to P.

5 Welfare

As we are considering a misspecified model, there are two notions of welfare we can study. First, the true welfare of the voter, that takes into account the true effect of wrong choices the voter may implement. Second, the welfare as perceived by the voter. In both cases, we compare welfare in “steady state”, i.e. when τ converges to the truth or remains in the cycle. Furthermore, we present results for the “good” state of the pool, where there si a fraction z of good politicians. We show that, even in this case, the long term welfare in the cycle may be higher than the long term welfare when τ converges to its true value.

5.1 True welfare

We can define the true welfare of the voter (“as if” she was a long living player) in a recursive form. To ease the notation, we state in the value functions only the identity of the incumbent (the challenger will be the opposite one).

5.1.1 True welfare in the cycle

Within the cycle $\tau_t = \tau_{t+1}$ and a T incumbent always delivers the good outcome. Under those conditions, the value function of having, at the beginning of time t, a T incumbent and a P challenger inherited from previous period while being within the cycle is

$$\begin{aligned} V_t(\tau_t, s_{I,t-1} = T) = {} & (1-\alpha(1-z\tau_t))[1+\beta V_{t+1}(\tau_t, s_{I,t} = T)]+ \\ & +\alpha(1-z\tau_t)[(1-\lambda^T)(1+\beta V_{t+1}(\tau_t, s_{I,t} = P, u_t = 1))+ \\ & +\lambda^T \beta V_{t+1}(\tau_t, s_{I,t} = P, u_t = 0)] \end{aligned} \tag{6}$$

This is because with probability $(1-\alpha(1-z\tau_t))$ ν adopts the "true" worldview, so T wins, delivers the good outcome and will be in power at the beginning of the next period. With probability $\alpha(1-z\tau_t)$ ν adopts the alternative worldview and hence P wins. In this case, we need to keep track of the outcome as well, because a populist in power producing a 0 outcome implies that a T will win the next election. Similarly, we have that the value function of having a P incumbent and a T challenger, when $u_{t-1}=1$ is

$$V_t(\tau_t, s_{I,t-1}=P, u_{t-1}=1) = \tag{7}$$
$$= (1-\lambda^T)(1+\beta V_{t+1}(\tau_t, s_{I,t}=P, u_t=1)) + \lambda^T \beta V_{t+1}(\tau_t, s_{I,t}=P, u_t=0)$$

Finally, the value function at the beginning of a period with a previous-period P incumbent producing a 0 outcome is

$$V_t(\tau_t, s_{I,t-1}=P, u_{t-1}=0) = 1 + \beta V_{t+1}(\tau_t, s_{I,t}=T) \tag{8}$$

because a T politician will surely win in period t. To ease the notation, we write $V_t(\tau_t, s_{I,t-1}=X, u_{t-1}) := V(\tau_t, X, u)$.

Note that in the steady state there is no updating on τ. As a consequence, it must be that $V_t(\tau,T) = V_{t+1}(\tau,T) := V(\tau,T)$ and $V_t(\tau,P,u) = V_{t+1}(\tau,P,u) := V(\tau,P,u)$ for every u. Imposing those equalities in (6), (7) and (8) and solving them jointly we find:

$$V(\tau,T) = \frac{1-\beta+\alpha(\beta-\lambda^T)(1-\tau z)+\beta\lambda^T}{(1-\beta)(1-\beta+\beta(\alpha(1-\tau z)+\lambda^T))} \tag{9}$$

$$V(\tau,P,1) = \frac{1-\beta+\alpha\beta(1-\lambda^T)(1-\tau z)+\lambda^T(2\beta-1)}{(1-\beta)(1-\beta+\beta(\alpha(1-\tau z)+\lambda^T))} \tag{10}$$

$$V(\tau,P,0) = \frac{1-\beta+\alpha\beta(1-\lambda^T)(1-\tau z)+\beta\lambda^T}{(1-\beta)(1-\beta+\beta(\alpha(1-\tau z)+\lambda^T))} \tag{11}$$

Note that we can re-write (9) as a weighted average of all the nodes following that state, weighted by their probability. In other words, we have

$$V(\tau,T) = (1-\alpha(1-\tau z)[1+\beta V(\tau,T)] + \alpha(1-\tau z)[(1-\lambda^T)(1+\beta V(\tau,P,1)) + \lambda^T \beta V(\tau,P,0)]$$

Those equations are different because they are measured at different point of the cycle. However, the limit for $\beta \to 1$ of the ratio of any couple of them gives always one.

5.1.2 True welfare when $\tau = 1$

We can calculate the steady state welfare in the good state (i.e. when $\gamma = z$, $\sigma = 1$ and τ approximates 1). This follows the same logic as above, as the P challenger can still win with positive probability, but everything is evaluated at $\tau = 1$ and $\sigma = 1$. The value function of having, at the beginning of time t, a T incumbent and a P challenger inherited from previous period while being in the "true" steady state is

$$\begin{aligned} V_t(1, s_{I,t-1} = T) &= (1 - \alpha(1-z))[\mu + \beta V_{t+1}(1, s_{I,t} = T)] + \\ &+ \alpha(1-z)[(1-\lambda^T)(1 + \beta V_{t+1}(1, s_{I,t} = P, u_t = 1)) + \\ &+ \lambda^T \beta V_{t+1}(1, s_{I,t} = P, u_t = 0)] \end{aligned} \tag{12}$$

This is because with probability $(1 - \alpha(1-z))$ ν adopts the "true" worldview, so T wins, delivers the outcome $\mu := z + (1-z)\lambda^T$ and will be in power at the beginning of the next period. With probability $\alpha(1-z)$ ν adopts the alternative worldview and hence P wins. In this case, we need to keep track of the outcome as well, because a populist in power producing a 0 outcome implies that a T will win the next election. Similarly, we have that the value function of having a P incumbent and a T challenger, when $u_{t-1} = 1$ is

$$\begin{aligned} &V_t(1, s_{I,t-1} = P, u_{t-1} = 1) = \\ &= (1-\lambda^T)(1 + \beta V_{t+1}(1, s_{I,t} = P, u_t = 1)) + \lambda^T \beta V_{t+1}(1, s_{I,t} = P, u_t = 0) \end{aligned} \tag{13}$$

Finally, the value function at the beginning of a period with a previous-period P incumbent producing a 0 outcome is

$$V_t(1, s_{I,t-1} = P, u_{t-1} = 0) = \mu + \beta V_{t+1}(1, s_{I,t} = T) \tag{14}$$

because a T politician will surely win in period t.

Solving the equations jointly we find:

$$V(1,T) = \frac{(1-\beta)\mu + \alpha(1-z)(1-\lambda^T - \mu(1-\beta)) + \beta\lambda^T\mu}{(1-\beta)(1-\beta+\beta(\alpha(1-z)+\lambda^T))} \tag{15}$$

$$V(1,P,1) = \frac{1-\beta+\alpha\beta(1-\lambda^T)(1-z)+\lambda^T(\beta(1+\mu)-1)}{(1-\beta)(1-\beta+\beta(\alpha(1-z)+\lambda^T))} \tag{16}$$

$$V(1,P,0) = \frac{(1-\beta)\mu+\alpha\beta(1-\lambda^T)(1-z)+\beta\lambda^T\mu}{(1-\beta)(1-\beta+\beta(\alpha(1-z)+\lambda^T))} \tag{17}$$

5.1.3 Comparison

First, we can study how those different steady state welfare vary with the parameters of the model, in the limit for $\beta \to 1$. To do so in a tractable way, we re-scale both welfare functions by $(1-\beta)$. More formally, we define:

$$V^\tau := lim_{\beta\to 1}(1-\beta)V(\tau,T) = \frac{\alpha(1-\lambda^T)(1-\tau z)+\lambda^T}{\alpha(1-\tau z)+\lambda^T} \tag{18}$$

and

$$V^1 := lim_{\beta\to 1}(1-\beta)W(1,T) = \frac{\alpha(1-\lambda^T)(1-z)+\lambda^T\mu}{\alpha(1-z)+\lambda^T} \tag{19}$$

The effect of α, λ^T and z on the true welfare of the voter in the two possible steady states that can be reached within the good pool (i.e. with $\gamma = z$) can be summarized as follows:

Proposition 5 *If α increases or z decreases, both V^τ and V^1 decrease. If λ^T increases, V^τ decreases and V^1 sometimes increases.*

The effects of α and z are quite obvious. In both cases, α raises the chances that an inefficient populist is elected, while z reduces those chances and also increases the performance of the T politician when there is no discipline (i.e. in W, where $\sigma = 1$). The effect of λ^T is different because it plays a different role. In the "populist cycle" an increase in λ^T means that the populist is more inefficient. In the "true" steady state there is an additional effect, given by the increase in μ: the higher is λ^T, the more efficient is the T politician (even when biased). Hence, the overall effect of λ^T can be positive.

Second, by comparing the limit, for $\beta \to 1$, of the ratio between (9) and (15) we can state the following:

Proposition 6 *If λ^T and z are sufficiently small, then $V^\tau > V^1$.*

Basically, proposition 6 states that, sometimes, the welfare in the populist cycle "steady state" (assuming parameters are such that it exists) can be above the welfare in the "good" steady state. This happens when $\lambda^T(\mu + 1) < 1$, i.e. when λ^T and z are sufficiently small. Intuitively, this comparison summarizes the basic trade off induced by the presence of populists: better behaviour from T politicians, but the country remains stuck in a cycle and elects suboptimal politicians with high probability ($\alpha(1 - \tau z)$ rather than $\alpha(1 - z)$). For this to be better than the "good" steady state, two things must happen. First, the good steady state must not be too good (i.e. z must be low, so that μ is small as well), otherwise there is no gain from discipline. Second, λ^T must be sufficiently small. An increase in λ^T is often good for the good steady state. On the other hand, its effect on the populist cycle is negative: a high λ^T implies that populists last in office for a shorter time, but also that they are more damaging when they are in office.

Figures 12 and 13 illustrates this point, showing that for low z and λ^T the populist cycle welfare (blue line) is actually higher than the one in the steady state of the good pool (orange line). This happens when the discipline effect is sufficiently strong and sufficiently important (i.e. z is small). If instead z is high, the cycle is always inferior than the good pool steady state. Figures 14 and 15 do the complementary exercise, fixing two different levels of λ^T and plotting the long term welfare as a function of z. When z is high, the welfare in $\tau = 1$ is always higher than the welfare in the cycle. However, the range of values of z where the opposite is true is bigger when λ^T is small.

Finally, note that Proposition 6 has an important implication: under some conditions, the behavioural assumption we are introducing is welfare improving for the voter. We state it more formally in the next corollary.

Corollary 1 *If the condition of Proposition 6 are met, then the long term welfare of the voter can be higher for some $\alpha > 0$ than for $\alpha = 0$.*

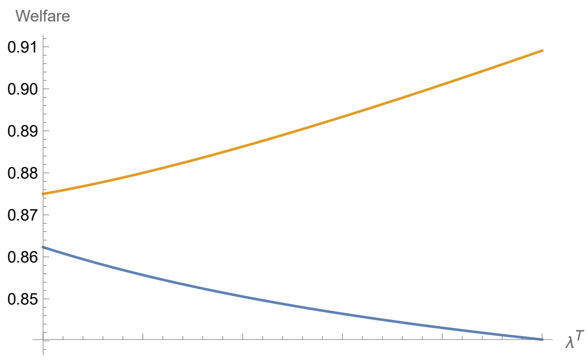

Figure 12: Comparison between steady state welfare with $\tau = 1$ (orange line) and steady state welfare in the cycle (blue line) as a function of λ^T for high z (i.e. 0.9).

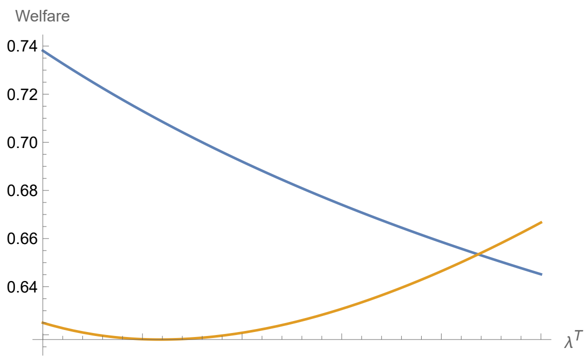

Figure 13: Comparison between steady state welfare with $\tau = 1$ (orange line) and steady state welfare in the cycle (blue line) as a function of λ^T for low z (i.e. 0.5).

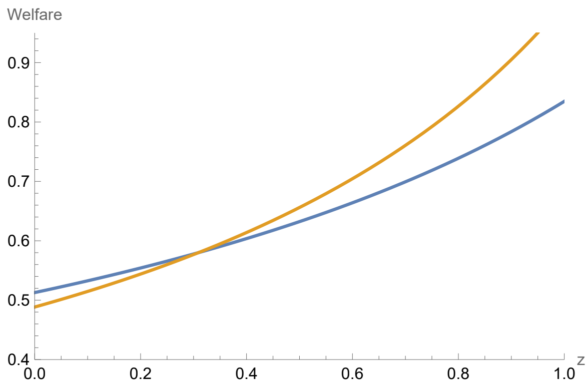

Figure 14: Comparison between steady state welfare with $\tau = 1$ (orange line) and steady state welfare in the cycle (blue line) as a function of z for high λ^T (i.e. 0.95).

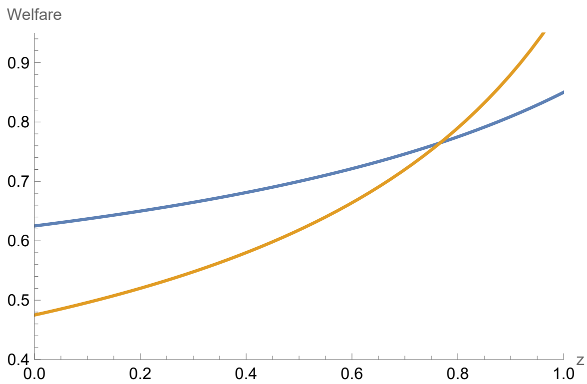

Figure 15: Comparison between steady state welfare with $\tau = 1$ (orange line) and steady state welfare in the cycle (blue line) as a function of z for low λ^T (i.e. 0.6).

The positive welfare effect of α goes trough the logic of Proposition 6, noticing that the stated condition holds for every α. When $\alpha = 0$, ν obtains a payoff of μ in every period in the steady state. If $\lambda^T(\mu + 1) < 1$, then being in the populist cycle is better. But the populist cycle requires a strictly positive (and sufficiently high) level of α.

5.2 Perceived welfare of the voter

Measuring the perceived welfare of the voter is not straightforward, as ν's perception of the probability distribution over the state of the world changes with its worldview adoption. One way to write it is assuming that ν knows she may adopt a different worldview, and then she measures the welfare conditional on that worldview (hence, for example, she wrongfully thinks that a P politician will always deliver a welfare of 1, because this is her best estimate given her beliefs that $\lambda = 0$). Hence, the perceived welfare, measure in the cycle and with a T incumbent and a P challenger, can be expressed recursively as

$$\tilde{V}(T) = \alpha(1 - \tau z)\frac{1}{1-\beta} + (1 - \alpha(1 - \tau z))(1 + \beta\tilde{V}(T)) \tag{20}$$

Equation (20) states that, with probability $\alpha(1 - \tau z)$, the voter adopts the populist worldview and hence she estimate to receive 1 forever. With the complement probability, she keeps the traditional worldview, she gets 1 in this period (as $\sigma = 0$ in the cycle) and then she chooses again in the next one, without any updating in τ. Re-arranging, we get that

$$\tilde{V}(T) = \frac{1}{(1-\beta)} \tag{21}$$

As above, we can now take the limit, for $\beta \to 1$, of the ratio between $\tilde{V}(T)$ and $W(1, T)$ or $V(\tau, T)$ respectively. Results are summarized in the following proposition:

Proposition 7 *In the limit for $\beta \to 1$, the perceived welfare in the cycle is higher both with respect to the steady state welfare when $\tau = 1$ and with respect to the true welfare of the cycle.*

Intuitively, ν's perceived welfare is always 1, because he over-estimates the effectiveness of the populist, when he adopts the populist worldviews. Hence, this is higher than both the other cases we consider. As a direct consequence, the voter always over-estimates her welfare in the cycle, compared to the welfare when τ converges to 1. To see this, note that the perceived welfare when $\tau \to 1$ is

$$\tilde{V}(T, \tau = 1) = \alpha(1 - \tau z)\frac{1}{1-\beta} + (1 - \alpha(1 - \tau z))((z + (1 - z)\lambda^T) + \beta\tilde{V}(T)) \tag{22}$$

because in this case $\sigma = 1$ and hence the per-period payoff of a T politician is $z+(1-z)\lambda^T$. The comparison between (22) and (20) shows that the perceived welfare of the voter is always higher in the "populist cycle" than in the case where τ converges to 1. Figure

16 illustrates this point, showing that $(1-\beta)\tilde{V}(T)$ (blue, dashed) is always greater than $(1-\beta)\tilde{V}(T, \tau = 1)$ (orange, dashed). Another point to be noted is that, in the case illustrated by the figure, the perception of the voter is factually wrong. The true welfare in the cycle, represented by the blue solid line, is in fact smaller than the true welfare for $\tau = 1$.

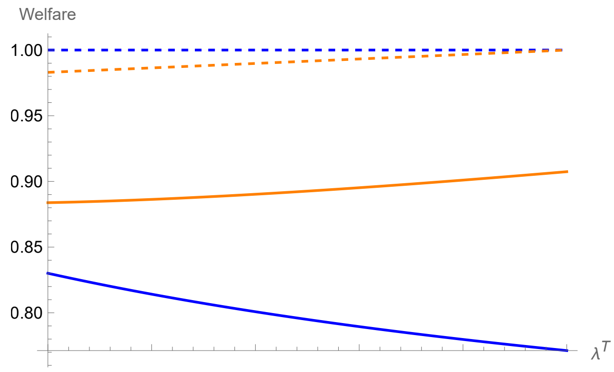

Figure 16: Comparison between normalized perceived welfare in the cycle (blue dashed) and normalized perceived welfare when $\tau = 1$ (orange, dashed) as a function of λ^T. Other parameters: $z = 0.9$, $\beta = 0.8$, $\alpha = 1$. We also plot the true (normalized) welfare functions when $\tau = 1$ (orange, solid) and in the cycle (blue, solid).

6 Conclusion

We developed a model in which distrust in the traditional political class can foster beliefs in simple alternative worldviews promoted by opportunistic populists. The main takeaways are the following. 1. When alternative facts are sufficiently pervasive, populist politicians, despite delivering (on average) an inferior outcome when compared with traditional politicians, can be a disciplining device for intermediate trust, improving the in office behaviour of biased traditional politicians. 2. When trust falls below a certain threshold, the presence of populist politicians generates political cycles where power alternates between traditional and populist politicians and political trust never recovers. 3. This cycle is not necessarily bad for voters' true welfare. 4. Welfare as perceived by the voters...systematically overestimate...much better than it is. 5. Overall, populist politicians and their simple alternative worldviews, always produce inferior outcomes in the short run, i.e., when in power, but they need not to be bad in the long run. When traditional politicians are (on average) sufficiently biased or the populist worldview is sufficiently close to the true one, welfare is higher in the cycle than in the steady state where trust converges to its true level.

Our result apply to the case of several western democracies, notably the US. The US has experienced a decline in public trust in recent years, and this has led to a pattern of cycles between Democrats and populist Republicans in elections, which appears likely to continue in the next years.

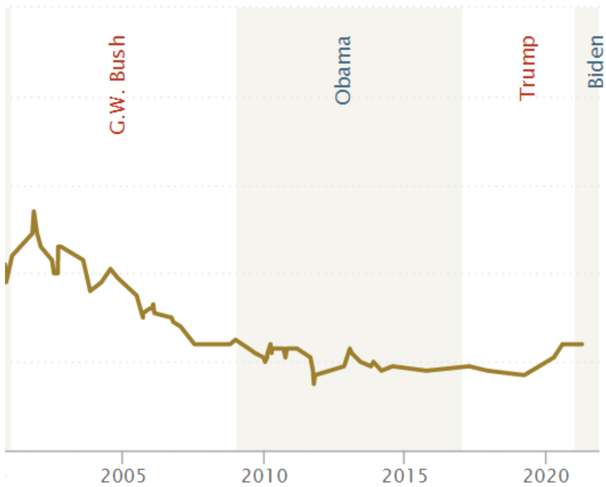

Figure 17: Public Trust in Government, PEW (2021)

Indeed, a key contributing factor for this is the perception that the Democratic Party has become too closely aligned with corporate interests and the political establishment, and has lost touch with the needs and concerns of working-class Americans. This has led to a perception that the party is out of touch with the concerns of everyday people, and has contributed to a growing sense of disillusionment with the political process. Another factor is the rise of populist rhetoric and anti-establishment sentiment, which has been embraced by many Republican candidates in recent years. This has been particularly effective in mobilizing support from disaffected voters who feel that the political system is rigged against them. These voters are the most likely to believe the alternative worldviews and narratives proposed by the MAGA side of the Republican party.

References

Andreottola, G. and E. Sartori (2023). *Simplistic Rhetoric and Poe's Law*. CSEF, Centre for Studies in Economics and Finance, Department of Economics ....

Ash, E., S. Mukand, and D. Rodrik (2021). Economic interests, worldviews, and identities: Theory and evidence on ideational politics. Technical report, National Bureau of Economic Research.

Ashworth, S. and E. Bueno De Mesquita (2014, August). Is Voter Competence Good for Voters?: Information, Rationality, and Democratic Performance. *American Political Science Review 108*(3), 565–587.

Auriol, E., N. Bonneton, and M. Polborn (2023). Shaking up the system: when populist disciplines elite politicians.

Becher, M., N. L. Marx, V. Pons, S. Brouard, M. Foucault, V. Galasso, E. Kerrouche, S. L. Alfonso, and D. Stegmueller (2021). Covid-19, government performance, and democracy: Survey experimental evidence from 12 countries. Technical report, National Bureau of Economic Research.

Bernhardt, D., S. Krasa, and M. Shadmehr (2022, October). Demagogues and the Economic Fragility of Democracies. *American Economic Review 112*(10), 3331–3366.

Blumenthal, B. (2022). Voter information and distributive politics.

Boffa, F., V. Mollisi, and G. A. Ponzetto (2023). Do incompetent politicians breed populist voters?: evidence from italian municipalities.

Bombardini, M., B. Li, and F. Trebbi (2023). Did us politicians expect the china shock? *American Economic Review 113*(1), 174–209.

Bordalo, P., N. Gennaioli, and A. Shleifer (2020). Memory, attention, and choice. *The Quarterly journal of economics 135*(3), 1399–1442.

Bordignon, M. and T. Colussi (2020). Dancing with the populist. new parties, electoral rules and italian municipal elections.

Bowen, T. R., D. Dmitriev, and S. Galperti (2023). Learning from shared news: When abundant information leads to belief polarization. *The Quarterly Journal of Economics 138*(2), 955–1000.

Campante, F., R. Durante, and F. Sobbrio (2018). Politics 2.0: The multifaceted effect of broadband internet on political participation. *Journal of the European Economic Association 16*(4), 1094–1136.

Carillo, M., D. Foarta, and M. Petrova (2023). Populism and complexity.

Crutzen, B. S. Y., D. Sisak, and O. H. Swank (2020). Left Behind Voters, Anti-Elitism and Popular Will. *SSRN Electronic Journal*.

Duggan, J. and C. Martinelli (2017). The political economy of dynamic elections: Accountability, commitment, and responsiveness. *Journal of Economic Literature 55*(3), 916–984.

Forand, J. G. (2021). Elections with job-motivated bureaucrats.

Gratton, G. and B. E. Lee (2023a). Drain the swamp: A theory of anti-elite populism.

Gratton, G. and B. E. Lee (2023b). Liberty, security, and accountability: The rise and fall of illiberal democracies. *Review of Economic Studies*, rdad030.

Guiso, L., H. Herrera, M. Morelli, and T. Sonno (2019, January). Global crises and populism: the role of Eurozone institutions*. *Economic Policy 34*(97), 95–139.

Guiso, L., H. Herrera, M. Morelli, and T. Sonno (2020). Economic insecurity and the demand of populism in europe. *Einaudi Institute for Economics and Finance*.

Guriev, S., N. Melnikov, and E. Zhuravskaya (2021). 3g internet and confidence in government. *The Quarterly Journal of Economics 136*(4), 2533–2613.

Guriev, S. and E. Papaioannou (2022, September). The Political Economy of Populism. *Journal of Economic Literature 60*(3), 753–832.

Levy, G. and R. Razin (2021). Short-term political memory and the inevitability of polarisation.

Levy, G., R. Razin, and A. Young (2022, March). Misspecified Politics and the Recurrence of Populism. *American Economic Review 112*(3), 928–962.

Little, A. T. (2019, July). The Distortion of Related Beliefs. *American Journal of Political Science 63*(3), 675–689.

Margalit, Y. (2019, November). Economic Insecurity and the Causes of Populism, Reconsidered. *Journal of Economic Perspectives 33*(4), 152–170.

Martinez-Bravo, M. and C. Sanz (2023, January). Trust and accountability in times of pandemic. Documentos de Trabajo 2302, Banco de España, Madrid. Series: Documentos de Trabajo.

Morelli, M., A. Nicolò, and P. Roberti (2021). A Commitment Theory of Populism. *SSRN Electronic Journal*.

Morelli, M. and G. Sasso (2021). Bureaucrats under populism. *Journal of Public Economics 202*, 104497.

Prato, C. and S. Wolton (2018). Rational ignorance, populism, and reform. *European Journal of Political Economy 55*, 119–135.

Silva, B. C. and C. Wratil (2023). Do parties' representation failures affect populist attitudes? evidence from a multinational survey experiment. *Political Science Research and Methods 11*(2), 347–362.

Svolik, M. W. (2013). Learning to love democracy: Electoral accountability and the success of democracy. *American Journal of Political Science 57*(3), 685–702.

Trombetta, F. (2020). When the light shines too much: Rational inattention and pandering. *Journal of Public Economic Theory 22*(1), 98–145.

Vosoughi, S., D. Roy, and S. Aral (2018). The spread of true and false news online. *science 359*(6380), 1146–1151.

A Proofs

Proof of Lemma 1. Note that if ν anticipates $\sigma = 0$ he is indifferent between P and T. We assume he breakes the indifference choosing P, as every other belief on σ would cause the choice to go in that direction. ∎

Proof of Lemma 2.

Recall that

$$\tau_{t+1}(u_t = 1) = \frac{\tau_t(1-(1-z)(1-\lambda^T)\sigma_t^c)}{1-(1-\lambda^T)\sigma_t^c(1-\tau_t z)}$$

$$\tau_{t+1}(u_t = 0) = \frac{\tau_t(1-z)\sigma_t^c}{\tau_t(1-z)\sigma_t^c + (1-\tau_t)\sigma_t^c} = \frac{\tau_t(1-z)}{1-\tau_t z}$$

Differentiating with respect to λ^T and z, while keeping σ_t^c fixed, we find the following:

$$\frac{\partial \tau_{t+1}(u_t = 1)}{\partial \lambda^T} \geq 0$$

$$(1-z)\tau_t\sigma_t(1-(1-\lambda^T)\sigma_t^c(1-\tau_t z)) \geq (1-\tau_t z)\sigma_t\tau_t(1-(1-z)(1-\lambda^T)\sigma_t^c)$$

$$1 - z \geq 1 - z\tau_t$$

which never holds. Hence, $\frac{\partial \tau_{t+1}(u_t=1)}{\partial \lambda^T} < 0$. Clearly, $\tau_{t+1}(u_t = 0)$ does not depend on λ^T. Moving to z, note that:

$$\frac{\partial \tau_{t+1}(u_t = 1)}{\partial z} \geq 0$$

$$(1-\lambda^T)\tau_t(1-(1-\lambda^T)\sigma_t^c(1-\tau_t z)) \geq \tau_t(1-\lambda^T)\tau_t(1-(1-z)(1-\lambda^T)\sigma_t^c)$$

$$1-(1-\lambda^T)\sigma_t^c(1-\tau_t z) \geq \tau_t(1-(1-z)(1-\lambda^T)\sigma_t^c)$$

$$1-\tau_t + (1-\lambda^T)\sigma_t^c((1-z)\tau_t - 1 + \tau_t z) \geq 0$$

$$1-\tau_t - (1-\lambda^T)\sigma_t^c(1-\tau_t) \geq 0$$

$$(1-\tau_t)(1-(1-\lambda^T)\sigma_t^c) \geq 0$$

that always holds. Finally, note that

$$\frac{\partial \tau_{t+1}(u_t = 0)}{\partial z} \geq 0$$
$$\tau_t(1-z) \geq 1 - \tau_t z$$

that never holds, hence $\frac{\partial \tau_{t+1}(u_t=0)}{\partial z} < 0$. ■

Proof of Proposition 1. Define $\Delta\tau_{t+1}(\sigma_t)$ as the difference in posteriors induced by the good and the bad outcome for any conjectured probability of distortion σ_t. Formally,

$$\begin{aligned}\Delta\tau_{t+1}(\sigma_t) &:= \tau_{t+1}(u_t = 1, \sigma_t) - \tau_{t+1}(u_t = 0, \sigma_t) \\ &= \frac{\tau_t(1-\tau_t)z}{(1-(1-\lambda^T)\sigma_t(1-\tau_t z))(1-\tau_t z)}\end{aligned}$$

First, note that $\Delta\tau_{t+1}(\sigma_t)$ is increasing in σ. To see this, note that $\tau_{t+1}(u_t = 0)$ is not a function of σ and that

$$\begin{aligned}\frac{\partial \tau_{t+1}(u_t = 1)}{\partial \sigma_t} &\propto -\tau(1-z)(1-\lambda^T)(1-(1-\lambda^T)\sigma_t(1-\tau z))+ \\ &+ (1-\lambda^T)(1-\tau z)\tau(1-(1-\lambda^T)(1-z)\sigma_t) \\ &= -(1-z) + 1 - \tau z \\ &= z(1-\tau) > 0\end{aligned}$$

Second, define $\tilde{\sigma}_t$ as the (unconstrained) solution of $\Delta\tau_{t+1}(\sigma_t)\alpha z E = 1$. We can find it in closed form, hence noticing that it is unique and continuous in τ. In particular, it is given by

$$\tilde{\sigma} = \frac{1 - \tau z(1 + (1-\tau)\alpha z E)}{(1-\lambda)(1-\tau z)^2}$$

Finally, note that in every equilibrium the conjectured σ must be equal to the actual one.

Hence, in order to have pure strategy equilibria, we need either $\Delta\tau_{t+1}(\sigma_t = 0)\alpha z E \geq 1$, meaning that the conjectured σ is 0 and the biased politician has an incentive to behave in the prescribed way when $\theta = 1$; or $\Delta\tau_{t+1}(\sigma_t = 1)\alpha z E \leq 1$, meaning that the bad T politician prefers to choose ful distortion and has the incentives to do so, when expected to behave in that way. This proves points 1 and 2.

Regarding point 3, note that we may have a $\tilde{\sigma}$ greater than 1 or smaller than 0. Consider the first case. If the value of σ that guarantees $\Delta\tau_{t+1}(\sigma)\alpha zE = 1$ is greater than 1, and given that $\Delta\tau_{t+1}(\sigma)$ is increasing in σ, it means that, when σ reaches its upper bound (i.e. 1), we have $\Delta\tau_{t+1}(\sigma)\alpha zE < 1$, hence a pure strategy equilibrium with full distortion. For a similar logic, if the value of σ that guarantees $\Delta\tau_{t+1}(\sigma)\alpha zE = 1$ is smaller than 0, and given that $\Delta\tau_{t+1}(\sigma)$ is increasing in σ, it means that, when σ reaches its lower bound (i.e. 0), we have $\Delta\tau_{t+1}(\sigma)\alpha zE > 1$, hence a pure strategy equilibrium with full distortion. Hence, rangers of parameters where $\tilde{\sigma}$ is above 1 or below 0 are already taken into account by the pure strategy equilibria described in 1 and 2. However, it is possible that $\tilde{\sigma} \in (0, 1)$. For those ranges of parameters (which may or may not exist) we have the mixed strategy equilibrium described by point 3. Note that all those equilibria are mutually exclusive (with the exception of the boundaries between them), that cover all possible parametric configurations and that there are no other PBNE with this form.

Finally, note that this equilibrium is unique among those where $u = 1$ does not cause a negative updating on τ. There may be other equilibria as follows:

- $\Delta\tau < 0$ and such that both players are indifferent when $\theta = 0$. In this case, we can have an equilibrium where the B politician chooses $x = 1$ when $\theta = 0$ and $x = 0$ when $\theta = 1$, and the G politician does the opposite, such that $Pr(u = 1|\gamma = 0) > Pr(u = 1|\gamma = z)$, so there is a positive updating observing $u = 0$.
- There may be an equilibirum where everyone chooses the wrong action so that $\hat{\tau}(u = 0) = \tau$ and $\hat{\tau}(u = 1)$ is off path (and with a $\tau = 0$ updating).

■

Proof of Proposition 2.

We look for α sufficiently small so that σ is always 1 in equilibrium. We drop the time subscript to ease the notation. In equilibrium, this happens when

$$1 \geq \Delta\tau\alpha z E$$

As $\Delta\tau$ is increasing in σ, we have that

$$\Delta\tau(\sigma) \leq \Delta\tau(\sigma = 1) = \frac{\tau(1-\tau)z}{(1-(1-\lambda^T)(1-\tau z))(1-\tau z)}$$

Hence, a sufficient condition for $1 \geq \Delta\tau\alpha z E$ is

$$\begin{aligned}
\alpha &\leq \frac{1}{\Delta\tau z E} \\
\alpha &\leq \frac{(1-(1-\lambda^T)(1-\tau z))(1-\tau z)}{\tau(1-\tau)zE} \\
\alpha &\leq \frac{(1-z)\lambda^T}{\frac{1}{4}zE} \\
\alpha &\leq \frac{4\lambda^T(1-z)}{zE}
\end{aligned}$$

where the third equality follows from the fact that $(1-z)\lambda^T < (1-(1-\lambda^T)(1-\tau z))(1-\tau z)$ for every τ and that $\tau(1-\tau)$ is maximized at $\frac{1}{4}$. The last inequality is just re-arranged. ■

Proof of Proposition 3.

When P is in government, there is no information transmission on τ, hence trivially $\tau_{t+1} = \tau_t$. When T is in government and $\tau_t < \bar{\tau}^0$, the $\sigma^* = 0$. As a consequence, $u_t = 1$ irrespective of the type of the T politician in charge, and $\tau_{t+1}(1) = \tau_t$. As τ does not move, the condition will remain true for $t+1, t+2, ...$ both with a T and a P politician in power. ■

Proof of Proposition 4.

Recall

$$Pr(PT) = \lambda^T$$

$$Pr(TP) = \alpha(1-\tau_{t+1}(u_t = 1)z) + \alpha(1-\gamma)(1-\lambda^T)\sigma_t z\Delta\tau \tag{A.1}$$

First part:

First, note that when $\sigma \in \{0,1\}$, neither $\Delta\tau$ nor $\tau_{t+1}(u_t = 1)$ are a function of α. As a consequence, it is trivial to note that the effect of α on $Pr(TP)$ is certainly positive.

Consider now the case of $\sigma = \frac{1-\tau_t z(1+(1-\tau_t)\alpha z E)}{(1-\lambda^T)(1-\tau_t z)^2}$. Substituting it into the equation for $\Delta\tau$ implies

$$\begin{aligned}
\Delta\tau &= \frac{\tau_t(1-\tau_t)z}{(1-\tau_t z) - (1-\lambda^T)(1-\tau_t z)^2 \frac{1-\tau z(1+(1-\tau_t)\alpha z E)}{(1-\lambda^T)(1-\tau_t z)^2}} \\
&= \frac{1}{\alpha z E}
\end{aligned}$$

Furthermore, $\tau_{t+1}(u_t = 1) = \frac{\tau_t}{1-\tau_t z} - \frac{\tau_t}{(1-\tau_t)\alpha z E}$. Substituting into (A.1), we obtain

$$Pr(TP) = \alpha\left(1 - z\tau_t\left(\frac{1}{1-\tau_t z} - \frac{1}{(1-\tau_t)\alpha z E}\right)\right) + \alpha(1-\gamma)(1-\lambda^T)z\frac{1-\tau_t z(1+(1-\tau_t)\alpha z E)}{(1-\lambda^T)(1-\tau_t z)^2}\frac{1}{\alpha z E}$$

This simplifies to

$$Pr(TP) = \alpha\left(1 - \frac{z\tau_t}{1-\tau_t z}\right) + \frac{\tau_t}{(1-\tau_t)E} + \frac{1-\gamma}{(1-\tau_t z)^2 E} - \frac{\tau_t z(1+(1-\tau_t)\alpha z E)(1-\gamma)}{(1-\tau_t z)^2 E} \tag{A.2}$$

Differentiating with respect to α and looking for conditions where the sign is negative, we find

$$\begin{aligned}
\frac{\partial Pr(TP)}{\partial\alpha} &< 0 \\
\frac{\tau_t z^2(1-\tau_t(1-\gamma))}{1-\tau_t z} &> 1 - 2\tau_t z \\
\tau_t z^2(1-\tau_t(1-\gamma)) &> (1-2\tau_t z)(1-\tau_t z) \\
\tau_t z((1-\gamma)z(1-\tau_t) + 3 - 2\tau_t z) &> 1
\end{aligned}$$

Second part:

The effect of λ^T on $Pr(PT)$ is obvious. Moving to $Pr(TP)$, note that

$$Pr(TP) = \alpha(1-\tau_{t+1}(u_t = 1)z) + \alpha(1-\gamma)(1-\lambda^T)\sigma_t z\left(\tau_{t+1}(u_t = 1) - \tau_{t+1}(u_t = 0)\right) \tag{A.3}$$

It is clear that, when parameters are such that $\sigma_t = 0$, $Pr(TP) = \alpha(1-\tau z)$ and as a consequence $\frac{\partial Pr(TP|\sigma_t^*=0)}{\partial \lambda^T} = 0$. Inspecting equation (A.2) reveals that also in the case of σ_t^* interior we have $\frac{\partial Pr(TP|\sigma_t^* \in (0,1))}{\partial \lambda^T} = 0$, because λ^T cancels out. Hence, it remains to consider only the case of σ_t^*. Note that, in this case,

$$\frac{\partial \tau_{t+1}(u_t = 1, \sigma_t^* = 1)}{\partial \lambda^T} = -\frac{\tau_t(1-\tau_t z)}{(1-(1-\lambda^T)(1-\tau_t z))^2}$$

Differentiating (A.3) with respect to λ^T while setting $\sigma_t = 1$ gives

$$\frac{\partial Pr(TP|\sigma_t^* = 1)}{\partial \lambda^T} = -\alpha z \frac{\partial \tau_{t+1}(u_t = 1, \sigma_t^* = 1)}{\partial \lambda^T} + \alpha(1-\gamma) z \left[-\Delta\tau + \frac{\partial \tau_{t+1}(u_t = 1, \sigma_t^* = 1)}{\partial \lambda^T}(1-\lambda^T) \right] \tag{A.4}$$

The RHS of (A.4) is negative iff

$$\begin{aligned}
&-\frac{\partial \tau_{t+1}(u_t = 1, \sigma_t^* = 1)}{\partial \lambda^T} - (1-\gamma)\Delta\tau + (1-\gamma)\frac{\partial \tau_{t+1}(u_t = 1, \sigma_t^* = 1)}{\partial \lambda^T}(1-\lambda^T) < 0 \\
&-\frac{\partial \tau_{t+1}(u_t = 1, \sigma_t^* = 1)}{\partial \lambda^T}(1-(1-\gamma)(1-\lambda^T)) < (1-\gamma)\Delta\tau \\
&\frac{\tau_t(1-\tau_t z)}{(1-(1-\lambda^T)(1-\tau_t z))^2}(1-(1-\gamma)(1-\lambda^T)) < (1-\gamma)\frac{\tau_t(1-\tau_t)z}{(1-(1-\lambda^T)(1-\tau_t z))(1-\tau_t z)} \\
&(1-\tau_t z)(1-(1-\gamma)(1-\lambda^T)) < (1-\gamma)(1-(1-\lambda^T)(1-\tau_t z)) \\
&1-\tau_t z < 1-\gamma \\
&\gamma < \tau_t z
\end{aligned}$$

As a consequence, it is clear that if $\gamma = 0$ then the condition holds, hence $\frac{\partial Pr(TP|\sigma_t^*=1)}{\partial \lambda^T} < 0$. If $\gamma = z$ the condition does not hold, hence $\frac{\partial Pr(TP|\sigma_t^*=1)}{\partial \lambda^T} > 0$. ■

Proof of Proposition 5. Recall that

$$V^\tau = \frac{\alpha(1-\lambda^T)(1-\tau z) + \lambda^T}{\alpha(1-\tau z) + \lambda^T}$$

and

$$V^1 = \frac{\alpha(1-\lambda^T)(1-z) + \lambda^T \mu}{\alpha(1-z) + \lambda^T}$$

with $\mu = z + (1-z)\lambda^T$. Moving to the comparative statics, and dropping the time subscript for analytical convenience:

$$
\begin{gathered}
\frac{\partial V^\tau}{\partial \alpha} \geq 0 \\
(1-\lambda^T)(\alpha(1-\tau z)+\lambda^T) \geq \alpha(1-\lambda^T)(1-\tau z)+\lambda^T \\
1-\lambda^T \geq 1
\end{gathered}
$$

which is clearly impossible. Therefore, $\frac{\partial V^\tau}{\partial \alpha} < 0$.

$$
\begin{aligned}
\frac{\partial V^\tau}{\partial \lambda^T} &\geq 0 \\
(-\alpha(1-\tau z)+1)(\alpha(1-\tau z)+\lambda^T) &\geq \alpha(1-\lambda^T)(1-\tau z)+\lambda^T \\
-\alpha^2(1-\tau z)^2-\alpha(1-\tau z)\lambda^T+\alpha(1-\tau z)+\lambda^T &\geq \alpha(1-\tau z)(1-\lambda^T)+\lambda^T \\
-\alpha(1-\tau z)-\lambda^T+1 &\geq 1-\lambda^T \\
-\alpha(1-\tau z) &\geq 0
\end{aligned}
$$

that holds only with equality if $\alpha = 0$. Therefore, unless $\alpha = 0$, $\frac{\partial V^\tau}{\partial \lambda^T} < 0$.

$$
\begin{gathered}
\frac{\partial V^\tau}{\partial z} \geq 0 \\
\alpha(1-\lambda^T)(1-\tau z)+\lambda^T \geq (1-\lambda^T)(\alpha(1-\tau z)+\lambda^T) \\
1 \geq 1-\lambda^T
\end{gathered}
$$

which is always true. Therefore, $\frac{\partial V^\tau}{\partial z} > 0$.

$$
\begin{aligned}
\frac{\partial V^1}{\partial \alpha} &\geq 0 \\
(1-\lambda^T)(\alpha(1-z)+\lambda^T) &\geq \alpha(1-\lambda^T)(1-z)+\lambda^T\mu \\
1-\lambda^T &\geq \mu \\
1-\lambda^T &\geq z+\lambda^T-\lambda^T z
\end{aligned}
$$

which never holds. To see this, note that the LHS is at most 0.5, while the RHS cannot be below 0.5, since $\lambda^T > 0.5$. Therefore, $\frac{\partial V^1}{\partial \alpha} < 0$.

$$\frac{\partial V^1}{\partial z} \geq 0$$
$$(-\alpha(1-\lambda^T) + \lambda^T(1-\lambda^T))(\alpha(1-z) + \lambda^T) + \alpha(\alpha(1-\lambda^T)(1-z) + \lambda^T\mu) \geq 0$$
$$-\alpha(1-\lambda^T)\lambda^T + (1-\lambda^T)\lambda^T\alpha(1-z) + (1-\lambda^T)(\lambda^T)^2 + \alpha\lambda^T\mu \geq 0$$
$$\alpha(z + (1-z)\lambda^T - (1-\lambda^T)z) + (1-\lambda^T)\lambda^T \geq 0$$
$$\alpha\lambda^T + (1-\lambda^T)\lambda^T \geq 0$$

which always holds. Therefore, $\frac{\partial V^1}{\partial z} \geq 0$. Finally,

$$\frac{\partial V^1}{\partial \lambda^T} \geq 0$$
$$(-\alpha(1-z) + \mu + \lambda^T(1-z))(\alpha(1-z) + \lambda^T) - \alpha(1-\lambda^T)(1-z) - \lambda^T\mu \geq 0$$
$$-\alpha^2(1-z)^2 - \alpha(1-z)\lambda^T + \left[\mu + (1-z)\lambda^T\right](\alpha(1-z) + \lambda^T) - \alpha(1-z)(1-\lambda^T) - \lambda^T\mu \geq 0$$
$$-\alpha^2(1-z)^2 - \alpha(1-z) + \left[\mu + (1-z)\lambda^T\right](\alpha(1-z) + \lambda^T) - \lambda^T\mu \geq 0$$
$$-\alpha^2(1-z)^2 - \alpha(1-z) + \mu\alpha(1-z) + \alpha(1-z)^2\lambda^T + (1-z)(\lambda^T)^2 \geq 0$$
$$\alpha(1-z)\lambda^T + \mu\alpha + (\lambda^T)^2 \geq \alpha(\alpha(1-z) + 1)$$
$$(\lambda^T)^2 \geq \alpha(1-z)(\alpha - (2\lambda^T - 1))$$

Inspecting the equation, it is clear that the LHS is strictly increasing in λ^T while the RHS is strictly decreasing in it. Hence, if the min of the LHS, i.e. $\frac{1}{4}$, is below the RHS evaluated in that point, i.e. $\alpha^2(1-z)$, then $\frac{\partial V^1}{\partial \lambda^T}$ is decreasing for sufficiently small λ^T and then increasing (note that the RHS is below 1). Otherwise, if $\frac{1}{4} \geq \alpha^2(1-z)$, $\frac{\partial V^1}{\partial \lambda^T} > 0$. ■

Proof of Proposition 6. Recall that

$$V^\tau = \frac{\alpha(1-\lambda^T)(1-\tau z) + \lambda^T}{\alpha(1-\tau z) + \lambda^T}$$

and

$$V^1 = \frac{\alpha(1-\lambda^T)(1-z) + \lambda^T\mu}{\alpha(1-z) + \lambda^T}$$

Then we have that

$$
\begin{aligned}
&\frac{V^\tau}{V^1} > 1\\
&\alpha^2(1-z)(1-\lambda^T)(1-\tau z) + \lambda^T\alpha(1-z) + (\lambda^T)^2 + \lambda^T\alpha(1-\lambda^T)(1-\tau z) >\\
&> \alpha^2(1-z)(1-\lambda^T)(1-\tau z) + \alpha(1-\tau z)\mu\lambda^T + \mu(\lambda^T)^2 + \alpha(1-\lambda^T)(1-z)\lambda^T\\
&\alpha(1-z)\lambda^T + \lambda^T(1-z)(1-\lambda^T) + \alpha(1-\tau z)(1-\lambda^T(\mu+1)) > 0
\end{aligned}
$$

As the first two terms of the last line are always positive, a sufficient condition for the inequality to hold is $\lambda^T(\mu+1) < 1$. ■

Proof of Corollary 1.

If the condition of Proposition 6 are met, we have that $V^\tau > V^1$. When $\alpha = 0$, only $V^1(\alpha = 0) = \mu$ is reached in the long term, because P is never chosen and $\sigma^* = 0$.

Consider now α sufficiently big so that the populist cycle can be reached. In this case, there exists a long term welfare $V^\tau = \frac{\alpha(1-\lambda^T)(1-\tau z)+\lambda^T}{\alpha(1-\tau z)+\lambda^T}$. The corollary is true if

$$
\begin{aligned}
V^\tau &> V^1(\alpha = 0)\\
\frac{\alpha(1-\lambda^T)(1-\tau z)+\lambda^T}{\alpha(1-\tau z)+\lambda^T} &> \mu\\
\alpha(1-\lambda^T)(1-\tau z)+\lambda^T &> (z+(1-z)\lambda^T)\alpha(1-\tau z) + (z+(1-z)\lambda^T)\lambda^T\\
\lambda^T(1-z(1-\lambda^T)-\lambda^T) &> \alpha(1-\tau z)(z(1-\lambda^T)+2\lambda^T-1)
\end{aligned}
$$

As we are looking for a sufficient condition, note that the RHS is strictly smaller than $(z(1-\lambda^T)+2\lambda^T-1)$ because $\alpha \leq 1$ and $\tau > 0$. Thus, a sufficient condition is

$$
\lambda^T(1-\lambda^T)(1-z) > z(1-\lambda^T) + 2\lambda^T - 1
$$

Note that the RHS is strictly increasing in z while the LHS is strictly decreasing in it. Furthermore, the RHS is above the LHS for $z = 1$. When $z = 0$, the LHS is above the RHS iff $-(\lambda^T)^2 - \lambda^T + 1 > 0$, i.e. iff $\lambda^T < \frac{-1+\sqrt{5}}{2}$. Therefore, if $\lambda^T < \frac{-1+\sqrt{5}}{2}$, we can always find a sufficiently sall z such that the condition is verified. ■

Proof of Proposition 7. As both V^τ and V^1 are strictly below 1, the limit for $\beta \to 1$ of the ratio between $\bar{V}$ and either $V(1, T)$ or $V(\tau, T)$ is always above 1. ■

B Extension: Two traditionals and one populist

In this section, we enrich the model assuming that there are three parties. Two of them are traditionals, and the third one is a populist. We maintain the same assumptions as the rest of the model.

B.1 Benchmark: $\alpha = 0$

If $\alpha = 0$, the populist party is irrelevant for the voter's decision. As a consequence, the race is effectively between the two T candidates. Note that the voter chooses at the beginning of every period and cannot commit to a future election strategy. As a consequence, the voter is effectively indifferent between the two T candidates, as they give her the exact same expected payoff. As they come from the same pool of politicians, and they live for one period only, there are not adverse selection issues in the voter-politician agency problem, but only moral hazard. Moreover, the game exhibits multiplicity of equilibria, because any election strategy is sequentially rational. Those strategies may or may not depend on the outcome of the previous period, and may or may not be able to solve the moral hazard problem. Different election strategies imply different equilibria (some of them are payoff equivalent, but not all of them). We will focus on time invariant equilibria, where ν's strategies are constant over time.

More formally, define $r(1)^T$ and $r(0)^T$ as the probabilities that the voter chooses the same T party that was in power in the previous period, as a function of the outcome of the previous period, u_{t-1}. The voter is forward looking and indifferent (in expectation) between the two candidates she is facing, so anything is acceptable. However, the way in which ν is expected to behave in the following period determines σ. We focus on symmetric, time invariant, full distortion and no distortion equilibria, but there can be other type of equilibria as well.

Lemma B1 *If ν's election strategies are such that $r(1)^T - r(0)^T > \frac{1}{E}$, the equilibrium exhibit no distortion. If instead $r(1)^T - r(0)^T < \frac{1}{E}$ the equilibrium exhibits full distortion.*

Proof of Lemma B1. In every symmetric, time-invariant equilibrium, the voter is indifferent between the two T candidates because they provide the same expected utility, ex ante. The incumbent politician anticipates the election strategies and chooses how to behave. As before, the only choice with a trade off is for the B politician when $\theta_t = 1$. In this case, by choosing $x_1 = 0$, she gets an expected payoff of $1 + Er(0)^T$. By choosing $x_1 = 1$ she gets $0 + Er(1)^T$. Therefore, it is an equilibrium to choose full distortion iff $0 + Er(1)^T \leq 1 + Er(0)^T$. Re-arranging, when election strategies are such that $r(1)^T - r(0)^T > \frac{1}{E}$, no distortion is the unique (symmetric, time invariant) equilibrium. When $r(1)^T - r(0)^T < \frac{1}{E}$, full distortion is the unique (symmetric, time invariant) equilibrium. In the knife-edge case that $r(1)^T - r(0)^T = \frac{1}{E}$ both those equilibria co-exists, together with any other interior σ. Finally, note that if $E < 1$ only full distortion equilibria exist. ■

Intuitively, if staying in power depends on the outcome in a sufficiently strong way and parties care about that, even biased T politicians behave well, choosing the correct action. Those equilibria maximize ν's welfare, but there is no commitment power on ν's side and so other equilibria are possible.

We can write the election strategies that give rise to no distortion equilibria as

$$r(1)^T \geq \frac{1}{E} + r(0)^T \tag{B.1}$$

We can also plot it in the $r(0)^T, r(1)^T$ space (Figure B1), noticing that we need $E > 1$ and that the lower bound of the condition is a line with slope = 1. Thus, there cannot be no distortion equilibria where $r(1)^T$ and $r(0)^T$ do not depend on u_{t-1}, i.e. along the 45 degrees line. In particular, we can focus on point $r(0)^T = 1, r(1)^T = 1$ (black dot, top right corner). This is the same election strategy that ν chooses when there is only one T (and one P) and $\alpha = 0$. T is always kept in power, and in fact the equilibrium has no discipline. With only one T and $\alpha = 0$, this is the unique equilibrium. With two T, instead, the voter can credibly condition te election strategies to the previous outcome, because there is an alternative (the other T candidate) that is as good as the T candidate of the party that was in power in the previous period. Thus, election strategies like $r(0)^T = 0, r(1)^T = 1$ (top right corner), able to guarantee no distortion, are also

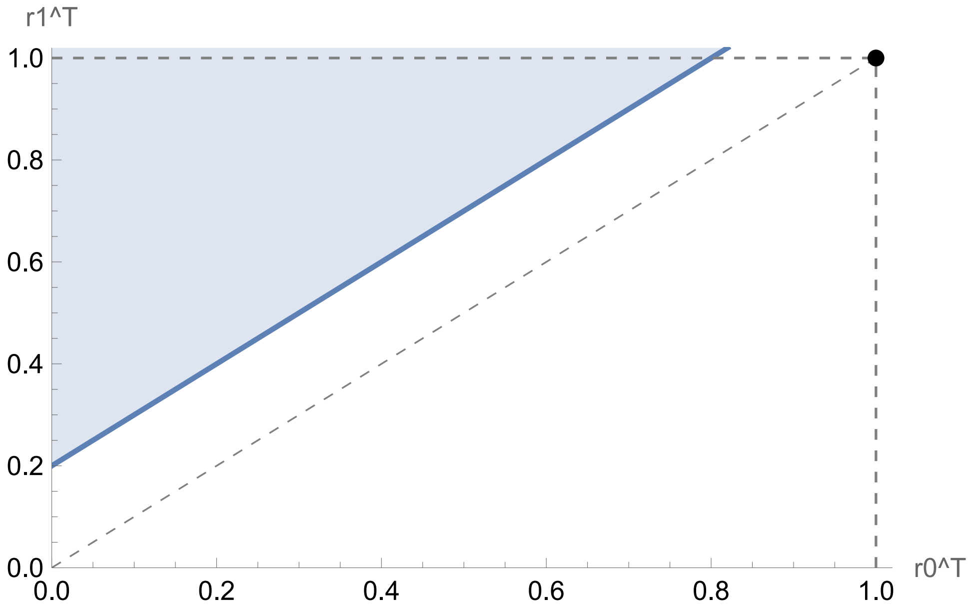

Figure B1: No distortion equilibria, $\alpha = 0$, $r(0)^T, r(1)^T$ space.

possible in equilibrium.

B.2 The role of α

Suppose now that α is positive. This implies that P's worldview may be chosen with positive probability. Now an incumbent producing outcome 1 knows that her party will remain in power with probability

$$r(1) = (1 - \alpha(1 - z\tau(1, \sigma^c)))r(1)^T$$

And, similarly, if the outcome is 0 the probability of remaining in power becomes

$$r(0) = (1 - \alpha(1 - z\tau(0, \sigma^c)))r(0)^T$$

In this case, of course, the updated τ is also a function of the conjectured σ, that must be correct in equilibrium. However, once the voter adopts the "traditional" worldview, she is once again indifferent between the two T candidates, so any election strategy is again possible in equilibrium, and different election strategies may give rise to different equilibria.

Lemma B2 *If election strategies are such that*

$$r(1)^T \geq \frac{1}{E}\frac{1}{1-\alpha+\alpha z\tau(1,\sigma=0)} + \frac{1-\alpha+\alpha z\tau(0,\sigma=0)}{1-\alpha+\alpha z\tau(1,\sigma=0)}r(0)^T \tag{B.2}$$

the equilibrium exhibits no distortion.

If election strategies are such that

$$r(1)^T \leq \frac{1}{E}\frac{1}{1-\alpha+\alpha z\tau(1,\sigma=1)} + \frac{1-\alpha+\alpha z\tau(0,\sigma=1)}{1-\alpha+\alpha z\tau(1,\sigma=1)}r(0)^T \tag{B.3}$$

the equilibrium exhibits full distortion.

Proof of Lemma B2. In every symmetric, time invariant equilibrium, ν is indifferent between the two T candidates. Therefore, if she adopts the T worldview, she can randomize between them with any probability, both conditional or unconditional to the outcome in the previous period. As before, the only choice with a trade off is for the B politician when $\theta_t = 1$. In this case, by choosing $x_1 = 0$, she gets an expected payoff of $1 + E(1-\alpha(1-z\tau(0,\sigma^c)))r(0)^T$. By choosing $x_1 = 1$ she gets $0 + E(1-\alpha(1-z\tau(1,\sigma^c)))r(1)^T$. In equilibrium, σ^c must be correct. Therefore, a full discipline equilibrium requires

$$(r(1)^T - r(0)^T)(1-\alpha) + \alpha z(\tau(1,\sigma=0)r(1)^T - \tau(0,\sigma=0)r(0)^T) \geq \frac{1}{E}$$

Re-arranging we get condition (B.2). A similar logic applies to full distortion equilibria. Finally, note that as $\tau(1,\sigma=0)$ is bigger than $\tau(1,\sigma=1)$, while $\tau(0)$ is independent of σ, the RHS of (B.2) is above the RHS of (B.3). If re-election strategies are in between the two conditions, we have an interior σ in equilibrium. ∎

Conditions (B.2) and (B.1) have a very similar structure. The difference is given by $\frac{1}{1-\alpha+\alpha z\tau(1,\sigma=0)}$, that multiplies $\frac{1}{E}$, and by $\frac{1-\alpha+\alpha z\tau(0,\sigma=0)}{1-\alpha+\alpha z\tau(1,\sigma=0)}$, that multiplies the slope. As $\frac{1}{1-\alpha+\alpha z\tau(1,\sigma=0)} > 1$ and $\frac{1-\alpha+\alpha z\tau(0,\sigma=0)}{1-\alpha+\alpha z\tau(1,\sigma=0)} < 1$, it is clear that condition (B.2) has an higher intercept and a smaller slope than (B.1).

Figure B2 adds condition (B.2) to Figure B1, capturing the two effect of the introduction of a P contender in a game with two traditionals. On the one hand, now there are no distortion equilibria even along the 45-degrees line, hence where the election strategy does not depend on u_{t-1}. In particular, we can look once again at the strategy $r(0)^T = 1, r(1)^T = 1$, i.e. the equilibrium election rule when there is only one T and the correct worldview is adopted. When α is positive, that point can be included (under some conditions) in the area with no distortion equilibria. This mirrors the result of the first part of the paper, where the introduction of a credible P challenger (i.e. a sufficiently high α) was able to discipline biased T politicians.

On the other hand, there are combinations of $r(1)^T$ and $r(0)^T$ that are no longer able to support a no distortion equilibrium. Intuitively, this happens because α has two competitive effects, in this environment. First, as in the benchmark model, it introduces the concern that P may win, hence the incentive not to push down τ. This effect implies that there can be equilibria with discipline where previously there was none, as in the baseline model. Second, it may reduce the dependence of the re-election from the outcome. In fact, with $\alpha = 0$ maybe it was enough to produce the good outcome to secure that the party remains in power. When α is positive, this is no longer the case, as P may still prevail. Hence, the disciplining effect of election strategies conditional on past outcomes is reduced.

B.3 Welfare

We now consider the effect of α on ν's true welfare, in this environment. First, consider the case of $\alpha = 0$. There, different equilibria produce different long term welfare. In equilibria where condition (B.1) is satisfied, there is no distortion (i.e. $\sigma^* = 0$) and also no updating on τ. Thus, the political system is in a steady state from the very beginning of the game. Furthermore, as $\alpha = 0$, P is never chosen. However, as there is no distortion, they are able to guarantee to the voter a payoff of 1 in every period. As a consequence, the net present value of the long term welfare in this case is

$$V^{TTP}(\alpha = 0, \sigma^* = 0) = \frac{1}{1-\beta}$$

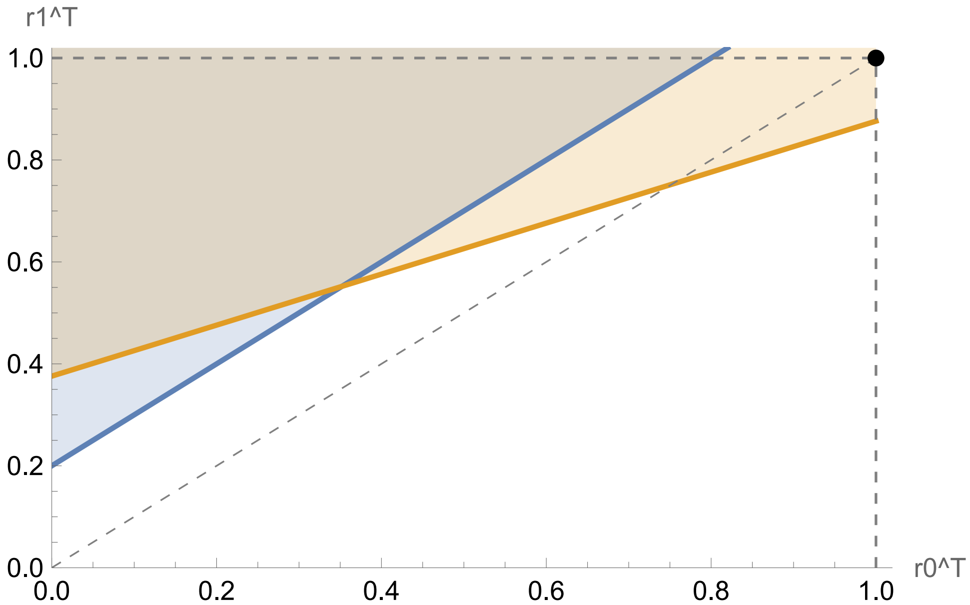

Figure B2: No distortion equilibria condition: $\alpha = 0$ (blue), $\alpha = 0.9$ (orange), $r(0)^T, r(1)^T$ space.

If, instead, election strategies are such that there is full distortion in equilibrium, τ in the long term converges to 1 and the net present value is

$$V^{TTP}(\alpha = 0, \sigma^* = 1) = \frac{\mu}{1 - \beta}$$

When α is positive, instead, the welfare is already captured by V^τ and V^1 (in the cycle and when $\tau = 1$ respectively) as defined in Section 5. In order to make menaningful comparisons, we study the effect of α on long term welfare keeping constant $r(0)^T$ and $r(1)^T$.

Proposition B1 *If the equilibrium is such that both T politicians choose $\sigma^* = 0$ irrespective of α, then the welfare effect of α is always negative.*

If the equilibrium is such that T politicians choose $\sigma^* = 1$ when $\alpha = 0$, then an increase in α can have a positive effect on welfare.

Proof of Proposition B1.

First part. Suppose the equilibrium election strategies $r(0)^T$ and $r(1)^T$ are such that $\sigma^* = 0$ for every α. Then, the true welfare of the voter is given by $\frac{1}{1-\beta}$. Any increase in α

implies that P is elected with positive probability (even with $\tau = 1$), bringing a per period payoff smaller than 1 and without changing the full discipline equilibrium. Therefore, the total present discounted value of the true welfare, given by $V^1(\sigma^* = 0) = \frac{\alpha(1-\lambda^T)(1-z)+\lambda^T}{\alpha(1-z)+\lambda^T}$, decreases in α.

Second part. Suppose the equilibrium election strategies $r(0)^T$ and $r(1)^T$ are such that $\sigma^* = 1$ when $\alpha = 0$. Over time, the udpating correctly converges to $\tau = 1$ and therefore the present discounted vaule of the steady state welfare is $V^1(\alpha = 0) = \mu$. If α increases and it becomes such that, for the same election strategies, the equilibrium shifts to a full discipline populist cycle, the true welfare in the cycle is now captured by $V^\tau = \frac{\alpha(1-\lambda^T)(1-\tau z)+\lambda^T}{\alpha(1-\tau z)+\lambda^T}$. As shown by Corollary 1, we can find parameters where $V^\tau > V^1(\alpha = 0)$. Therefore, a sufficiently big increase in α can have a positive welfare effect. ■

To get an intuition, consider for example the equilibrium with $r(0)^T = 0, r(1)^T = 1$. In that case, if $E > 1$, we have no distortion with $\alpha = 0$. If parameters are such that $\frac{1}{E}\frac{1}{1-\alpha+\alpha z\tau} < 1$, we have no distortion also when α is positive. However, raising α implies that the voter moves from a long term welfare able to guarantee 1 in every period, to a situation where P is elected with positive probability, thus causing the per-period welfare to drop below 1, sometimes. This cost comes without discipline gains on the traditional side, because traditionals are already disciplined.

The second part of Proposition B1 basically follows from Proposition 6. Consider as an example the equilibrium where $r(0)^T = 1, r(1)^T = 1$. In this case, there is full distortion when $\alpha = 0$. In the long run, τ goes to 1 and the voter obtains μ in every period. However, if there exists α sufficiently high so that condition (B.2) is met, then the polity may instead reach a populist cycle where τ does not move anymore, P is elected but also T are disciplined. As shown in section 5 and in Proposition 6, under some conditions the net present value of voter's welfare in the cycle is higher than $\frac{\mu}{1-\beta}$. Hence, an increase in α can have a positive welfare effect.

Overall, Proposition B1 provides an important qualification to the welfare effect of the populist cycle. It is positive (relative to the case of no populists, or of $\alpha = 0$) if the political competition between two traditional parties is not able to solve the moral hazard problem, i.e. it is not able to discipline biased T politicians. When this is the case, having a populist as a credible alternative and ending up in a populist cycle may be welfare enhancing. If, instead, the political system is such that biased T are disciplined, then increasing α can only be detrimental to voters' welfare. Furthermore, those equilibria may co-exists. This implies that, even when the fundamentals of different polities are the identical, the disciplining role (or lack of thereof) of voters is crucial.

C Extension: Endogenous populist entry

In this extension, we allow for endogenous populist entry. In other words, we start with a world in which there is only one traditional party and we study under what conditions a populist competitor chooses to enter the political arena. Note that we get similar results also in a setup where we have two parties and one of them can turn populist.

C.1 Timing with entry

The timing is now as follows. At the beginning of the game, Nature chooses the true quality of the pool of politicians and τ_0. Then, in every period:

1. State variables: τ_t, $s_{I,t-1}$, $s_{C,t-1}$ (where the latter are the identities of the incumbent and the challenger, if there is one. If the challenger is not active, we define $s_{C,t-1} = \emptyset$);
2. ν chooses the worldview λ_t and votes;
3. Entry decision by the perspective populist challenger, upon paying a cost ξ;
4. The winning politician observes $s_{C,t}$, learns Γ^i, θ and chooses x_t (constrained to 1 if $s = P$);
5. u_t is observed and ν calculates the posterior $\hat{\tau}_t$;
6. Period ends;

C.2 Within period analysis

The post-elections equilibrium behaviour of every period with a T and a P politician available is as described above. We need to discuss the case of one T politician alone.

C.2.1 One T politician

Suppose there is no (populist) challenger. Then, the T politician knows her party will remain in power tomorrow irrespective of her policy choice. As a consequence, in the unique equilibrium there is full distortion, i.e. $\sigma^* = 1$.

Lemma C1 *Suppose only the T party is active. Then, in the unique equilibrium,* $\sigma^* = 1$.

Proof of Lemma C1. If there is no challenger, a B traditional politician faces no trade off, hence $\sigma^* = 1$ is the unique NE. ■

C.2.2 Identity choice

Given this result, and the equilibrium described in the main body of the paper, we can now move back to the identity choice in period t, immediately after the elections, and the subsequent (on path) policy choice. First, note that the potential entrant compares the expected utility of running as a P, taking into account how this affects the probability of victory for her party in $t+1$, with the cost of entry. Furthermore, note that - when she chooses - she has the same information as ν, because she does not know her individual type. The expected utility of running as a as a P is

$$\mathbb{E}(v(T)|\tau_t, \lambda^T, \lambda_t) = \mathbb{E}(\alpha(1 - \tau_{t+1}z)E|\tau_t, \lambda^T, \lambda_t) \tag{C.1}$$

where the expectation is taken over everything the incumbent may do in period t: the state, her individual type, the probability of policy distortion. What matters in this case, of course, is $\mathbb{E}(\tau_{t+1}|\tau_t, \lambda^T, \lambda_t)$, i.e. the expected ν's beliefs at the beginning of period $t+1$ (or, equivalently, at the end of period t). Differently from ν, the politician does not adopt alternative worldviews. However, when the incumbent is a T, the challenger updating works exactly as the voter's updating, because by Lemma 1 a voter electing a T holds the correct worldview. As a consequence, by the martingale property of Basyesian updating,

$$\mathbb{E}(\tau_{t+1}|\tau_t, \lambda^T = \lambda_t) = \tau_t \tag{C.2}$$

Comparing (C.1) and ξ, and using (C.2), we can derive the following result.

Lemma C2 *In every period t in which the incumbent has a T identity, we have a populist challenger active iff* $\tau_t \leq \frac{\alpha E - \xi}{z\alpha E} := \tau^P$.

Proof of Lemma C2. When choosing whether to enter, the party compares $\mathbb{E}(v(T)|\tau_t, \lambda^T, \lambda_t) = \alpha(1 - \tau_t z)E$ with the cost ξ. Rearranging for τ gives $\tau_t \leq \frac{\alpha E - \xi}{z\alpha E}$. ■

Lemma C2 is quite useful, as it defines precisely a threshold, in terms of τ_t, that determines whether we have a populist party or not. A necessary condition for this to happen is a sufficiently low level of trust in politicians. This happens because, by committing to a worldview and the policy consistent with it, populists "shut down" the risk of a T politician captured by special interests. When this probability is high, a populist becomes an attractive alternative. Second, if $z < \frac{\alpha E - \xi}{\alpha E}$, then the challenger would pick a P identity for every τ_t, hence he would always be there. This implies that, when the "baseline" level of trust in the pool of politicians is sufficiently low, populism is more likely to arise. Finally, if $\alpha E < \xi$ the populist challenger never runs. Hence, if alternative worldviews are not sufficiently powerful, there is no point in using them, irrespective of the level of trust. A direct consequence is that, if we set $\alpha = 0$, i.e. we assume that distrust in politicians does not translate into a higher probability of "buying" the vision of the world proposed by P, there is no purpose for the challenger to run as P, as her worldview will not be adopted and her party would surely lose the elections.

Furthermore, we can derive some important comparative statics on τ^P.

Lemma C3 *τ^P is decreasing in z and ξ and increasing in α and E.*

Proof of Lemma C3. It follows directly from straightforward differentiation of $\tau^P = \frac{\alpha E - \xi}{z\alpha E}$. ■

All the results of lemma C3 make intuitive sense. It says that populist entry is more likely in polities with a low level of baseline trust and with a high penetration of alternative worldviews.

To complete this section, we need to characterize the on path policy choice, noticing that they depend on whether there is a T incumbent facing a P challenger or two T parties. Proposition C1 summarizes it.

Proposition C1 *There exists a PBNE of the game such that:*

- *As long as the P challenger stays out, there is full distortion ($\sigma_t = 1$);*

- *If one of the parties has a P identity, σ_t behaves as described in Lemma 1;*
- *The challenger runs with a P identity iff $\tau_t \leq \tau^P$*

Proof of Proposition C1. Follows directly from the argument in the text. ■

As a consequence, we can outline the on-path σ_t (for the equilibria selected above) for a generic period starting with τ_t provided that the winner at the start-of-the period elections is a T. We do so in figure C1, blue line. When $\tau_t > \tau^P$ the challenger remains out and the equilibrium σ_t is 1. Below that level, the challenger runs as a P and as a consequence the equilibrium σ_t follows the orange dotted line (i.e. the equilibrium choice of a T incumbent facing a P challenger). Finally, the green dotted line is the equilibrium σ_t if the challenger remains out.

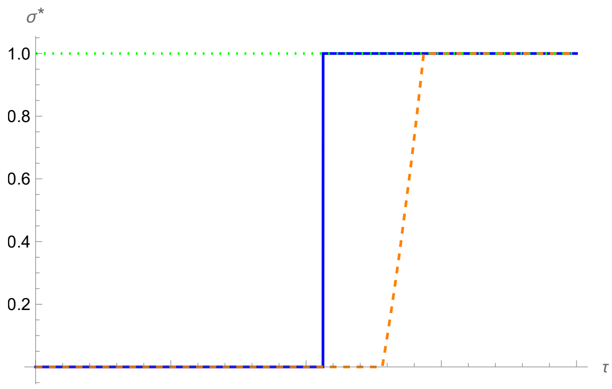

Figure C1: Equilibrium unconditional probability of distortion as a function of τ_t (blue solid line); conditional on a T challenger (green dotted line); conditional on a P challenger (orange, dashed line). Other parameters: $E = 4$, $\xi = 3.3$, $\alpha = 1$,$z = 0.8$, $\lambda^T = 0.6$. We plot the interval $\tau_t \in (0.6, 1)$.

C.3 Dynamics and cycles

The dynamics, with α high enough, is similar to the one described above. The main difference is given by the role plaid by τ^P, as it determines the threshold where the system switches from a T politician alone to a T vs P case. For tractability we assume that the populist party has to pay a cost ξ in every period in order to keep running when out of office. Hence, if τ_t moves back above τ^P, the P party temporarily leaves the arena. In this section, we assume parametric restrictions to focus on parameters range where

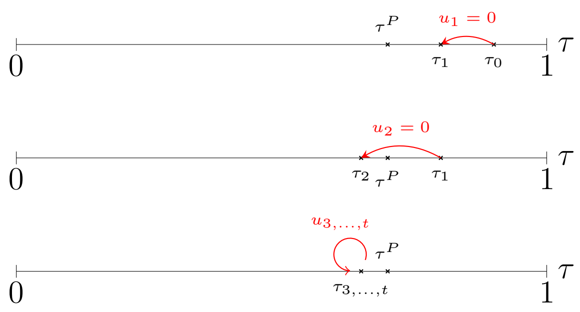

Figure C2: One possible path for the evolution of τ from τ_0 to the cycle. When τ_t moves below τ^P, i.e. after two consecutive negative outcomes, the challenger becomes a populist and trust remains fixed at τ_2 forever.

results are interesting and the model is tractable. Hence, we assume that E is sufficiently big so that there is a nonempty set of τ_t such that $\sigma_t = 0$. Furthermore, we assume that $\tau^P < \bar{\tau}^0$ and that $\tau_0 > \tau^P$. The first assumption is basically guaranteed by a sufficiently big E, as long as $\tau^P \in (0,1)$, as shown in Appendix E. One important point to notice is that $\tau_{t+1}(u_t = 1) = \tau_t$ when $\sigma_t = 0$, meaning that there cannot be updating on trust without some positive probability of distortion. We are now in the position to assess the dynamic evolution of the system.

Suppose $\tau_0 > \tau^P$. This implies that $s_{I,0} = T$ and $s_{C,0} = \emptyset$ because the challenger is inactive. Furthermore, $\sigma_0 = 1$. If the incumbent politician is of B type and $\theta_0 = 1$, the outcome will be bad, i.e. $u_0 = 0$. As a consequence, $\tau_1 < \tau_0$. Obviously, if $\theta_1 = 0$ then trust move upward, and so on. Suppose however that $\tau_1 > \tau^P$. Again, $s_{C,1} = T$, $\sigma_1 = 1$ and, if $\theta_1 = 0$ and the incumbent politician is of B type, $u_1 = 0$. Hence, $\tau_2 < \tau_1$. If this process continues, there is a positive probability to reach $\tau_t < \tau^P$. As soon as this happens, we have that $s_{C,t} = P$ and $\sigma_t = 0$. As a consequence, $u_t = 1$ for every θ_t, but $\tau_{t+1} = \tau_t$. This, combined with the fact that τ_t does not move when the incumbent is a populist, implies that this condition is absorbing: irrespective on who is in power, $\tau_{t+1} = \tau_t$; T incumbents behave well, even if biased, but trust is not restored; the incumbent switches from T to P with probability τ_t and from P to T with probability λ^T. A P challenger remains such even after losing, as he may gain power again in the next period (with probability τ_t). We call this situation a "cycling" regime, whose path of entry is described in Figure C2.

Note that a direct consequence of the cycle is the fact that the system does not converge to the true state. Once populists kick in, trust cannot be restored endogenously.

C.4 Simulations

To get some insights on this dynamic, we simulate the behaviour of τ over time. All the graphs that follow, we set $\tau_0 = 0.9$, $E = 4$, $\xi = 3.35$, $z = 0.8$ and $\lambda^T = 0.6$. We run 200 simulations each 80 periods long, conditional to the good state of the political pool.

Figure C3 shows that, over time, trust moves toward the true state (i.e. $\tau = 1$) when no populists are allowed and $\sigma = 1$. If we allow for endogenous switch to a P platform, we see two different scenarios. On the one hand, trust may move toward the truth. However, still we may enter in the "cycling" regime with positive probability. τ tends to move upward, but there may be enough negative realizations that push it below τ^P. Then, it stays there, as shown in figure C4.

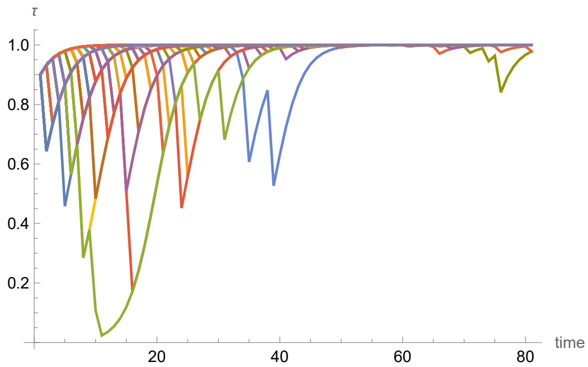

Figure C3: Evolution of τ over time. No P allowed.

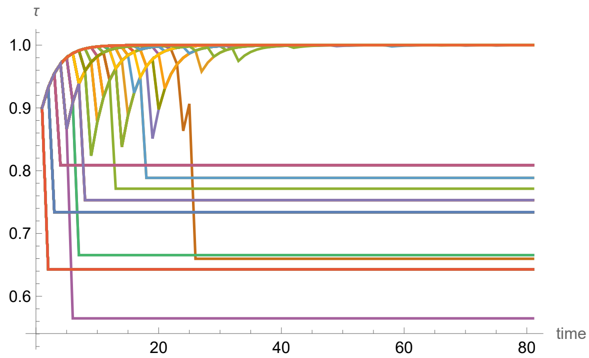

Figure C4: Evolution of τ over time. Endogenous P.

C.5 True welfare of the voter

Our results on the true welfare of the voter are also robust to this extension of the model. The true welfare in the cycle behaves exactly in the same way, while voter's welfare when τ converges to 1 is now higher: the biased T politician chooses full distortion, but now there is no populist (as long as parameters are such that $\tau^P < 1$), hence there is zero risk for the voter to elect an inefficient politician. In particular, the voter's welfare in the good state (i.e. when $\gamma = z$ and τ approximates 1) is

$$\bar{V}(\tau = 1) = \frac{z + \lambda^T(1 - z)}{(1 - \beta)} \tag{C.3}$$

By comparing the limit, for $\beta \to 1$, of the ratio between (9) and (C.3) we can state the following:

Proposition C2 *It is possible to find sufficiently small values of z and λ^T such that the limit for $\beta \to 1$ of $\frac{V(\tau,T)}{\bar{V}(\tau=1)} > 1$.*

Proof. Recall that

$$V^\tau = \frac{\alpha(1 - \lambda^T)(1 - \tau z) + \lambda^T}{\alpha(1 - \tau z) + \lambda^T}$$

and

$$\bar{V}^1 = (1 - \beta)\bar{V}(\tau = 1) = z + \lambda^T(1 - z) := \mu$$

The limit for $\beta \to 1$ of $\frac{V(\tau,T)}{\bar{V}(\tau=1)}$ is greater than 1 iff

$$\begin{aligned} \alpha(1 - \lambda^T)(1 - \tau z) + \lambda^T &> \mu(\alpha(1 - \tau z) + \lambda^T) \\ \lambda^T(1 - \mu) &> \alpha(1 - \tau z)(\lambda^T - (1 - z)(1 - \lambda^T) \\ \lambda^T(1 - z)(1 - \lambda^T) &> \alpha(1 - \tau z)(\lambda^T - (1 - z)(1 - \lambda^T) \end{aligned}$$

A sufficient condition for the last inequality to hold is

$$\begin{aligned} \lambda^T(1 - z)(1 - \lambda^T) &> \lambda^T - (1 - z)(1 - \lambda^T \\ (1 - z)(1 - \lambda^T)(\lambda^T + 1) &> \lambda^T \end{aligned}$$

Note that the LHS is strictly decreasing in z, while the RHS does not depend on it. Furthermore, the condition never holds for $\lambda^T = 1$ and $z = 1$. Hence, it is possible to find a threshold in z sufficiently small to ensure that the condition holds as long as $(1 - \lambda^T)(\lambda^T + 1) > \lambda^T$. This last inequality is true for $\lambda^T < \frac{-1+\sqrt{5}}{2}$. ■

Basically, proposition C2 confirms that, sometimes, the welfare in the populist cycle can be above the welfare in the "good" steady state even when identity is endogenous.

This happens when λ^T and z are sufficiently small.

Finally, note that Proposition C2 has an important implication: under some conditions, the behavioural assumption we are introducing is welfare improving for the voter. More broadly, an increase in α can be (steady-state) welfare improving, as summarized by the next corollary.

Corollary C1 *If parameters are such that Proposition C2 holds, then an increase in α can increase the expected steady state welfare.*

Proof of Corollary C1. Consider parameters where Proposition C2 holds for every α. If α is very small, over time τ converges to its true value, and this guarantees a steady state welfare of $(1-\beta)\mu$. If α increases up to the point where the cycle exists and hence it is reached in the long run with positive probability, then the expected steady state welfare will be a convex combination between $(1-\beta)\mu$ and the welfare in the cycle, which is higher than just $(1-\beta)\mu$ because of proposition C2. ■

Intuitively, consider a situation where Proposition C2 holds. Suppose, however, that $\alpha < \bar{\alpha}$, hence over time τ necessarily converges to the truth (i.e. to $\tau = 1$): in this case, long term welfare converges to $\bar{W}$. An increase in α that allows for the cycle in equilibrium, instead, introduces another possible steady state, that guarantees a higher utility, thus increasing the expected steady state welfare.

D Extension: Different worldview acquisition process

In the main body of the paper, we assumed that the evolution process of the acquired worldview depends on the identity of the incumbent, namely that ν goes back to the T worldview upon observing a populist failure, but chooses with the trust-dependent probability when the incumbent is a T politician. We now consider the alternative scenario where, at the beginning of every period, ν chooses the P worldview with probability $\alpha(1-z\tau_t)$ and chooses the T one otherwise, irrespective of the identity of the incumbent.

In this case, nothing changes in terms of incentives for T politicians or in terms of equilibrium structure in general (including the cycle, as there is still no updating with a P in power). The main difference is in terms of steady state welfare.

D.1 True welfare in the cycle

Within the cycle $\tau_t = \tau_{t+1}$ and a T incumbent always delivers the good outcome. Under those conditions, the value function of having, at the beginning of time t, a T incumbent and a P challenger inherited from previous period while being within the cycle is

$$\begin{aligned} V_t(\tau_t, s_{I,t-1} = T) = (1 - \alpha(1 - z\tau_t))[1 + \beta V_{t+1}(\tau_t, s_{I,t} = T)] + \\ + \alpha(1 - z\tau_t)(1 - \lambda^T + \beta V_{t+1}(\tau_t, s_{I,t} = P)) \end{aligned} \tag{D.1}$$

This is because with probability $(1-\alpha(1-z\tau_t))$ ν adopts the “true” worldview, so T wins, delivers the good outcome and will be in power at the beginning of the next period. With probability $\alpha(1 - z\tau_t)$ ν adopts the alternative worldview and hence P wins. Similarly, we have that the value function of having a P incumbent and a T challenger is

$$\begin{aligned} V_t(\tau_t, s_{I,t-1} = P) = (1 - \alpha(1 - z\tau_t))[1 + \beta V_{t+1}(\tau_t, s_{I,t} = T)] + \\ + \alpha(1 - z\tau_t)(1 - \lambda^T + \beta V_{t+1}(\tau_t, s_{I,t} = P)) \end{aligned} \tag{D.2}$$

Note that in the steady state there is no updating on τ. As a consequence, it must be that $V_t(\tau, T) = V_{t+1}(\tau, T) := V(\tau, T)$ and $V_t(\tau, P) = V_{t+1}(\tau, P) := V(\tau, P)$. Imposing those equalities in (D.1) and (D.2) solving them jointly we find:

$$V(\tau, T) = V(\tau, P) = \frac{1 - \alpha(1 - \tau z)\lambda^T}{1 - \beta} \tag{D.3}$$

D.2 True welfare when $\tau = 1$

We can calculate the steady state welfare in the good state (i.e. when $\gamma = z$, $\sigma = 1$ and τ approximates 1). This follows the same logic as above, as the P challenger can still win with positive probability, but everything is evaluated at $\tau = 1$ and $\sigma = 1$.

$$\begin{aligned} V_t(1, s_{I,t-1} = T) = (1 - \alpha(1 - z))[\mu + \beta V_{t+1}(1, s_{I,t} = T)] + \\ + \alpha(1 - z)(1 - \lambda^T + \beta V_{t+1}(1, s_{I,t} = P)) \end{aligned} \tag{D.4}$$

This is because with probability $(1 - \alpha(1 - z))$ ν adopts the "true" worldview, so T wins, delivers the outcome $\mu := z + (1 - z)\lambda^T$ and will be in power at the beginning of the next period. With probability $\alpha(1 - z)$ ν adopts the alternative worldview and hence P wins. In this case, we need to keep track of the outcome as well, because a populist in power producing a 0 outcome implies that a T will win the next election. Similarly, we have that the value function of having a P incumbent and a T challenger is

$$\begin{aligned} V_t(1, s_{I,t-1} = P) = (1 - \alpha(1 - z))[\mu + \beta V_{t+1}(1, s_{I,t} = T)] + \\ + \alpha(1 - z)(1 - \lambda^T + \beta V_{t+1}(1, s_{I,t} = P)) \end{aligned} \tag{D.5}$$

Solving the equations jointly we find:

$$V(1, T) = V(1, P) = \frac{(1 - \alpha(1 - z))\mu + \alpha(1 - z)(1 - \lambda^T)}{1 - \beta} \tag{D.6}$$

D.3 Comparison

Comparing (D.3) and (D.6), it is clear that the usual trade-off between better behaviour and election of an inefficient populist is at work in this case as well. In fact, on the one hand $\mu < 1$, capturing the fact that in the cycle the T politician behaves better. On the other hand, still $\mu > 1 - \lambda^T$ and $\alpha(1-z) \leq \alpha(1-z\tau)$, capturing the fact that, in the cycle, the election of an inefficient P is a more likely outcome.

Proposition D1 *If* $(1-z)(1-\lambda^T) \geq \lambda^T(1-\tau)$ *then welfare is higher in the cycle than in the steady state.*

Proof of Proposition D1. The ratio between (D.3) and (D.6) is above 1 when

$$
\begin{aligned}
1-\alpha(1-\tau z)\lambda^T &> (1-\alpha(1-z))\mu + \alpha(1-z)(1-\lambda^T) \\
1-\alpha(1-\tau z)\lambda^T &> \lambda^T + (1-\lambda^T)z - \alpha(1-z)(\lambda^T - (1-\lambda^T)(1-z)) \\
(1-\lambda^T)(1-z) - \lambda^T(\alpha(1-\tau z) - \alpha(1-z)) &> \alpha(1-z)(1-\lambda^T)(1-z) \\
(1-\lambda^T)(1-z)(1-\alpha(1-z)) &> \lambda^T\alpha z(1-\tau)
\end{aligned}
$$

As we are looking for a sufficient condition, and the LHS is decreasing in α while the RHS is increasing in α, we find when the condition is satisfied for $\alpha = 1$. Substituting, we find $(1-\lambda^T)(1-z) > \lambda^T(1-\tau)$. ■

Proposition D1 formalizes the intuition above, outlining a sufficient condition for welfare being higher in the cycle. As expected, this condition is more likely to be met for small z and small λ^T, hence when discipline is important (small z) and when the populist worldview is not too far from reality (small λ^T).

E Restrictions on E

Recall that there are values of τ where $\sigma = 0$ iff $E > \frac{2-z+2\sqrt{1-z}}{\alpha z^2}$.

To see this, note that an equilibrium with $\sigma = 0$ requires that, at least for some values of τ, we have that $1 \leq \frac{\tau(1-\tau)z}{1-\tau z} z\alpha E$. Re arranging for τ, the inequality becomes

$$\tau^2 z^2 \alpha E - \tau z(1 + z\alpha E) + 1 \leq 0$$

For this to happen in the range $(0, 1)$ we need two conditions. First, it must be that

$$-\frac{-z(1 + z\alpha E)}{2z^2\alpha E} \in (0, 1)$$

The only thing to check is the upper bound, that is true as long as

$$\begin{aligned} &1 + z\alpha E < 2z\alpha E \\ &E > \frac{1}{z\alpha} \end{aligned}$$

Second, it must be that

$$\begin{aligned} &z^2(1 + z\alpha E)^2 > 4z^2\alpha E \\ &1 + z^2\alpha^2 E^2 + 2z\alpha E > 4\alpha E \\ &z^2\alpha^2 E^2 - 2\alpha(2 - z)E + 1 > 0 \end{aligned}$$

This requires

$$\begin{aligned} E &\geq \frac{2\alpha(2-z) + \sqrt{4\alpha^2(2-z)^2 - 4z^2\alpha^2}}{2z^2\alpha^2} \\ E &\geq \frac{2\alpha(2-z) + 2\alpha\sqrt{(2-z)^2 - z^2}}{2z^2\alpha^2} \\ E &\geq \frac{(2-z) + 2\sqrt{1-z}}{z^2\alpha} \end{aligned}$$

Finally, note that this second condition is more restrictive than the first one, as

$$\begin{aligned}
\frac{(2-z)+2\sqrt{1-z}}{z^2\alpha} &> \frac{1}{z\alpha} \\
\frac{(2-z)+2\sqrt{1-z}}{z} &> 1 \\
2-2z+2\sqrt{1-z} &> 0
\end{aligned}$$

To complete the proof, note that the second root of the quadratic equation in E above, i.e. $E \leq \frac{(2-z)-2\sqrt{1-z}}{z^2\alpha}$, is incompatible with the first condition. In fact, we can show that

$$\begin{aligned}
\frac{(2-z)-2\sqrt{1-z}}{z^2\alpha} &< \frac{1}{z\alpha} \\
2(1-z) &< 2\sqrt{1-z} \\
\sqrt{1-z}\sqrt{1-z} &< \sqrt{1-z} \\
\sqrt{1-z} &< 1 \\
z &> 0
\end{aligned}$$

Solving for the values of τ that delimit the segment where it is an equilibrium to have $\sigma = 0$, we find $\underline{\tau}^0 = \frac{1+z\alpha E-\sqrt{(1+z\alpha E)^2-4\alpha E}}{2z\alpha E}$ and $\bar{\tau}^0 = \frac{1+z\alpha E+\sqrt{(1+z\alpha E)^2-4\alpha E}}{2z\alpha E}$. Note that the legth of the segment of τ where $\sigma = 0$ is an equilibrium is defined as

$$\begin{aligned}
\Delta\tau^0 &= \bar{\tau}^0 - \underline{\tau}^0 \\
&= \frac{\sqrt{(1+z\alpha E)^2-4\alpha E}}{z\alpha E} \\
&= \sqrt{\frac{(1+z\alpha E)^2-4\alpha E}{(z\alpha E)^2}} \\
&= \sqrt{\left(\frac{1}{z\alpha E}+1\right)^2 - \frac{4}{z\alpha E}}
\end{aligned}$$

We now show that the length of the interval is always increasing in E. Taking the derivative with respect to E, first note that $sign\left(\frac{\partial\Delta\tau^0}{\partial E}\right) = sign\left(\frac{\partial}{\partial E}\left(\left(\frac{1}{z\alpha E}+1\right)^2-\frac{4}{z\alpha E}\right)\right)$. Finally,

$$\begin{aligned}
\frac{\partial}{\partial E}\left(\left(\frac{1}{z\alpha E}+1\right)^2-\frac{4}{z\alpha E}\right) &= 2\left(\frac{1}{z\alpha E}+1\right)\left(-\frac{z\alpha}{(z\alpha E)^2}\right)+\frac{4z\alpha}{(z\alpha E)^2} \\
&\propto -\left(\frac{1}{z\alpha E}+1\right)+2 \\
&= \frac{2z\alpha E-1-z\alpha E}{z\alpha E} \\
&= \frac{z\alpha E-1}{z\alpha E}
\end{aligned}$$

The sign is positive, hence the interval is increasing in E, iff $z\alpha E > 1$. This is always the case when the interval exists, i.e. when E satisfies the assumption above. To see this, note that

$$\begin{aligned}
z\alpha E &\geq z\alpha\frac{2-z+2\sqrt{1-z}}{\alpha z^2} \\
&= \frac{2-z+2\sqrt{1-z}}{z} > 1
\end{aligned}$$

To see the last inequality, note that it implies $2-z+2\sqrt{1-z} > z$, that simplifies to $2(1-z)+2\sqrt{1-z} > 0$, which is always satisfied.

F Additional results

Derivation of Equation (3)

$$\begin{aligned}
\hat{\tau}(k,n) &= \frac{\tau_0(z(1-\lambda^T)+\lambda^T)^k((1-z)(1-\lambda^T))^{n-k}}{\tau_0(z(1-\lambda^T)+\lambda^T)^k((1-z)(1-\lambda^T))^{n-k}+(1-\tau_0)(\lambda^T)^k(1-\lambda^T)^{n-k}} \\
&= \left(1+\frac{1-\tau_0}{\tau_0}\left(\frac{\lambda^T(1-z)}{z(1-\lambda^T)+\lambda^T}\right)^k\left(\frac{1}{1-z}\right)^n\right)^{-1}
\end{aligned}$$

Derivation of Equation (5)

$$
\begin{aligned}
Pr(TP) &= (1-\gamma)(1-\lambda^T)\sigma_t[\alpha(1-\tau_{t+1}(u_t=0)z)] + (1-(1-\gamma)(1-\lambda^T)\sigma_t)[\alpha(1-\tau_{t+1}(u_t=1)z)] \\
&= \alpha(1-\tau_{t+1}(u_t=1)z) + \alpha(1-\gamma)(1-\lambda^T)\sigma_t z\Delta\tau
\end{aligned}
$$