A Dynamic Model of Political Narratives and the Rise of Neoliberalism*

Timothy Besley Adam Brzezinski

LSE LSE

March 25, 2026

Abstract

We develop a dynamic model where two parties compete for office by offering policies supported by narratives that describe policy effects. Voters demand narratives that suit their interests based on motivated reasoning and parties supply narratives to win elections. Over time, the set of narratives can expand endogenously. We characterize equilibrium paths and show how polarizing narratives can benefit a party through enabling policy differentiation. We then develop a specific application to "neoliberalism", a policy narrative that gained currency in the late 1970s. Our analysis clarifies why the 1970s oil shocks incentivized parties on the right to supply neoliberal narratives and why that prompted parties on the left to respond with the "Third Way" narrative that became a new consensus. These narrative dynamics were precipitated by a temporary shock but permanently altered the policy equilibrium, preventing a return to the old post-war consensus.

Keywords: narratives, motivated reasoning, neoliberalism
JEL Codes: D72, D91, H11

*We are grateful to Roland Bénabou, Jonathan Bendor, Steve Callander, Nathan Canen, Paul Collier, James Forder, Matt Gentzkow, Larry Kramer, Gilat Levy, Itzhak Rasooly, Andrei Shleifer, Steven Teles, David Tuckett, Stephane Wolton and a number of seminar participants for helpful comments and insights. We acknowledge financial support from the Programme for Cohesive Capitalism at the LSE and the British Academy. The paper previously had the title "The Political Economy of Neoliberal Narratives".

1 Introduction

Narratives that are used for interpreting how the world works have long been seen as important in shaping policy debates. Discussions of economic policy problems are frequently communicated to citizens via narratives as a means to influence their voting behavior and thereby to affect policy choices. Yet, until recently, policy narratives were largely ignored by theoretical frameworks in political economy despite increasing interest in their role among economists (e.g., Shiller, 2017).

But what is a narrative? We follow an emerging literature that conceptualizes narratives as subjective causal models (e.g., Eliaz and Spiegler, 2020). While this definition of narratives has only recently gained traction in economic theory, the core ideas have been explored using different terms such as subjective "models" (Schwartzstein and Sunderam, 2021; Levy et al., 2022; Montiel Olea et al., 2022), "ideology" (Bénabou, 2008; Izzo et al., 2023), or "mindset" (Stantcheva, 2024; Chinoy et al., 2023). Whatever the terminology, allowing for the idea that individuals can have their own conception of causal relationships provides new insights into how policy is determined in democratic societies.

This paper develops a dynamic model of political competition where policy change is driven by an endogenously expanding set of narratives. To do so, we build on the familiar two-party model of political competition with probabilistic voting (e.g., Persson and Tabellini, 2002). We add the possibility that electorally-motivated parties introduce narratives in an attempt to win elections. Voters demand narratives to make sense of the world and to inform their policy preferences. They look at the set of available narratives, adopting the one that they find most appealing, based on motivated reasoning. The narratives that voters adopt influence which policies are ex-post credible.

In political equilibrium, politicians will deploy new narratives only if they generate policy convictions that improve their electoral chances, which shapes the dynamic coevolution of narratives and policies. Our analysis characterizes the shape of different possible equilibrium paths, highlighting the importance of *polarizing narratives*, i.e. those that underpin policy differentiation across parties. We also provide conditions for *policy hysteresis* effects driven by narratives, where temporary economic changes can unleash a narrative dynamic that permanently alters equilibrium policies.

The focus of our model is to explore a new dimension of policy disagreement driven by narrative polarization, where groups of citizens endogenously choose competing causal models of how policy affects economic outcomes. This enriches

political disagreement beyond standard economic interests and allows us to capture Keynes' famous view expressed in the concluding paragraph of his General Theory of Employment, Interest, and Money (Keynes, 1936) that "the power of vested interests is vastly exaggerated compared with the gradual encroachment of ideas... sooner or later, it is ideas, not vested interests, which are dangerous for good or evil." And it chimes with the views of narrative psychologists that the human mind absorbs narratives to supplement logic and evidence-based thinking (e.g., Johnson et al., 2020).

Our model allows us to investigate narrative polarization in a structured way, driven by both supply and demand. This differs from standard models of political competition where voters are assumed to have a common understanding of how the economy works, usually underpinned by a reading of evidence that is assumed to be commonly accessible and objectively processed. In contrast to this, we assume that office-motivated politicians strategically supply narratives that combine causal parameters across the different dimensions in which policy affects outcomes into a coherent world view. Citizens then *choose* the narrative that maximizes their expected utility, a (subconscious) process of motivated reasoning. Put simply, citizens believe things because they are convenient rather than because they are evidence-based, an idea that has long been accepted by psychologists (see Kunda, 1990), but is relatively new in models within economics (Bénabou, 2015; Bénabou and Tirole, 2016) or political science (Little, 2019; Little et al., 2022). This approach allows us to investigate situations where citizens disagree because they use different subjective causal models rather than only because they have access to different evidence or have different interests.

After analyzing general features of the framework, we illustrate its utility in a specific context by discussing a narrative that emphasizes the virtues of free markets, often referred to as *neoliberalism*. This term is frequently invoked rhetorically in policy debates, mostly in a pejorative way. However, it did not originate as a pejorative term; indeed, some of its leading economic proponents proudly self-identified as neoliberals (Burgin, 2012). Among these was Milton Friedman, who wrote that "neo-liberalism offers a real hope of a better future, a hope that is already a strong cross-current of opinion and that is capable of capturing the enthusiasm of men of good-will everywhere, and thereby becoming the major current of opinion" (Friedman, 1951, p. 4). A contemporary example of self-identification with neoliberalism is the Adam Smith Institute in the UK, which in 2016 started to use the term to describe its policy views.[1]

[1]See `https://www.adamsmith.org/blog/coming-out-as-neoliberals`.

We refrain from taking a normative stance for or against policies referred to as "neoliberal". Instead, we analyze the conditions under which neoliberal narratives thrive in political competition. This, we argue, depends on two factors: whether neoliberal narratives are demanded by voters, and whether they are supplied by parties. On the demand-side, neoliberal narratives will only be popular if they give hope of a better future to (some) voters, as suggested in the quote by Friedman above. On the supply side, neoliberal narratives will be deployed by parties if that leads to an electoral advantage, which only happens under specific conditions.

In order to investigate the political success of neoliberalism, our framework gives a precise definition of the term. Specifically, neoliberal narratives are causal policy models with two key features: (i) they highlight the *trickle-down* benefits of policies which harm poorer individuals, arguing that losses will be more than offset by gains accruing from a larger size of the economy; and (ii) they question the ability of government to produce public goods efficiently.

The intellectual history of neoliberal thought stretches back to von Mises, Hayek, and other Austrian economists, amid the fall of the Austro-Hungarian empire after WWI (Slobodian, 2018; Burgin, 2012). It became common currency with the rise of the Chicago School, notably under the influence of Milton Friedman. Yet, for the most part, these ideas remained on the fringe of policy thinking, until they became influential following the oil crises of the 1970s and the subsequent rise to power of Margaret Thatcher and Ronald Reagan (Stedman Jones, 2014).

The rise of what is now called neoliberalism fractured a policy consensus that had emerged following the Great Depression and World War II, one that argued for a wider role for government in the form of demand management and regulation of markets. This *post-war consensus* travels under a variety of names. In the US, Gerstle (2022) refers to it as "the New Deal Order". In the UK, it is associated with the term "Butskellism", coined by *The Economist* magazine in 1954 to describe the convergence of views across the Labour and Conservative parties (see Harrison, 1999). Under this consensus, policies of parties from the left and right frequently resembled each other. In the US, Eisenhower, a Republican president, embraced most of the policy platforms of Franklin D. Roosevelt. In the UK, the Conservative party switched to supporting the National Health Service in the same time period. While there were dissenters in both cases, policy continuity remained the norm even as political administrations turned over.

That the rise of Margaret Thatcher and Ronald Reagan came with new policy narratives is hard to dispute. An iconic example of the shift in emphasis can be found in Ronald Reagan's inaugural address in 1981 (Reagan, 1981):

> We invented the assembly line and mass production, but punitive tax policies and excessive and unnecessary regulations plus government borrowing have stifled our ability to update plant and equipment. When capital investment is made, it's too often for some unproductive alterations demanded by government to meet various of its regulations.

Reagan's shift away from interventionism was a clear departure from views that had dominated policy-thinking in preceding decades. Moreover, the associated policies gained support from voters with both Reagan and Thatcher going on to be re-elected, in the latter case twice. Neoliberalism also fared well elsewhere in the same time period and the "Washington Consensus" became an established blueprint for policy reforms around the world.

To explain the success of Thatcher's and Reagan's narratives, an existing historical literature argues that the 1970s oil shocks were a catalyst for policy shifts away from the post-war consensus on high taxes and strong regulation (for example, Daunton, 2022). Our framework enriches our understanding of this by articulating the symbiotic relationship between policy *and* narrative shifts that underpinned this dynamic. It thereby resolves an apparent puzzle that, once there was a return to economic stability, there was no policy reversal. Policy hysteresis came in the form of the "Third Way", a hybrid counter-narrative that adopted elements of neoliberalism which became optimal to deploy after neoliberal thinking had emerged and led to a new policy consensus. Once new narratives have become part of the political discourse, they remain even in benign economic conditions, preventing a return to the previous policy equilibrium.

Related Literature Our model builds on a developing literature in political economy that investigates competition in narratives.[2] There is now an emerging convention in economic theory that defines narratives as subjective causal models.[3] However, the way in which competition in narratives is modeled, and how narratives are assumed to relate to available data, varies widely. For example, Eliaz and Spiegler (2020) assume that narratives may incorrectly specify the causal chain between variables, but that, conditional on that causal chain, beliefs fit available data (see also Eliaz et al., 2022). Schwartzstein and Sunderam (2021) posit that the most persuasive causal models are those that most closely fit the data to be explained, an idea also taken up by Aina (2023). Montiel Olea et al. (2022) study

[2]This literature complements the widespread approach of Bayesian persuasion models (Kamenica and Gentzkow, 2011; Kamenica, 2019).

[3]However, some papers define narratives differently (e.g., Bénabou et al., 2018).

agents who exogenously differ in what they consider to be potentially relevant explanatory variables of an outcome, and evaluate which type of agent would be most confident of their prediction against commonly observable data. In political economy, Izzo et al. (2023) offer a model where parties trade off offering causal models that better fit the data with the goal of influencing voter policy preferences (see also Montiel Olea and Prat, 2025). And Levy et al. (2022) develop a model where voters have exogenously given narratives, some of which are incorrect; conditional on these narratives, voters update beliefs consistently with the available data.

We suggest a way of modeling narrative selection which complements these existing approaches. Instead of supposing that it is the fit to the data that pins down a unique narrative for each type of agent, we suppose that there is a *set* of narratives that is acceptable to citizens on the basis of evidence so that parties have some leeway in narrative supply. Within this set, equilibrium narratives are determined through parties' strategic deployment of narratives to win elections in the face of citizens' willingness to adopt them based on motivated reasoning.

For many policy contexts, data may limit the set of plausible causal models but does not uniquely pin down one such model, creating a space for narratives to have influence. This reflects the fact that policy experts routinely disagree about how to read empirical evidence about the consequences of policy actions, sometimes even when using the same data source (Breznau et al., 2022). In principle, Bayesian citizens could assign different probabilities to each piece of evidence and aggregate the information; in politics, the incentives to do so for citizens who cannot plausibly influence an election are very limited (see also Caplan, 2008). Even if there were a unique data set and empirical approach that everyone agreed on, wide confidence intervals or lack of point-identification create a range of plausible narratives.

Throughout our analysis we emphasize the role of motivated reasoning in narrative adoption. This connects to an emerging literature in political economy that incorporates motivated reasoning in a range of applications (Bénabou and Tirole, 2016; Little, 2019; Little et al., 2022; Brzezinski, 2023). In our setting, motivated reasoning can help to explain why, even if a narrative is available, it may not be deployed in political equilibrium and why it may persist in spite of a stock of emerging evidence. The idea of motivated reasoning connects to an older tradition in economic theory that questions Bayesian belief formation (see Ortoleva, 2022, for a recent review). Previous approaches have highlighted the role of emotions against outside groups (Glaeser, 2005; Alesina and Stantcheva, 2020), social net-

works and group identity (Murphy and Shleifer, 2004; Shayo, 2009; Bonomi et al., 2021), and ideologies (Bénabou, 2008; Izzo et al., 2023). Allowing for motivated reasoning rather than evidence-based beliefs allows us to capture a source of policy persistence which would otherwise be hard to understand.

Our application adds a bridge between the formal study of narratives and more specific applications in political economy. We develop a model that is grounded in standard political economy models of competition between a left-wing and a right-wing party (Persson and Tabellini, 2002). This helps to make clear what the value added is of introducing narratives in studying policy formation in elections. In our application, we shed light on why narratives and policies developed from a post-war consensus in the 1950s towards polarization with the rise of neoliberal narratives in the 1980s and thence to a new Third Way consensus in the 1990s.[4]

Although our main purpose is to build a theoretical framework, we are closely allied to an emerging empirical literature that investigates narratives and related concepts. Several tools have been developed to analyze text in political speeches (Gentzkow et al., 2019; Ash et al., 2024). Moreover, there is increasing recognition that beliefs are not solely data-driven, but can be influenced by plot lines in soap operas (La Ferrara et al., 2012), reality shows (Kearney and Levine, 2015), divergent narratives in opinion programs on TV (Bursztyn et al., 2023b), videos of immigrant life experience (Alesina et al., 2023), justifications that give cover to dissenting opinions (Bursztyn et al., 2023a), light entertainment broadcasting (Durante et al., 2019), and plays (Hoff et al., 2021).

The remainder of the paper is organized as follows. Section 2 introduces our core model, which builds on standard approaches in political economy, but introduces the concept of narratives as subjective causal models of how policy affects outcomes. Section 3 offers general insights into possible equilibrium paths, highlighting the importance of polarizing narratives. Section 4 outlines a specific application of our model to neoliberal narratives. It then studies the dynamic evolution of deregulation narratives, shedding light on why neoliberalism supplanted the post-war consensus, and how it later developed into a new "Third Way" consensus narrative. We summarize our insights and directions for future research in section 5. Appendix A contains the proofs while Supplemental Appendices B and C contain additional material.

[4]This dynamic approach complements existing static models that try to understand free-market ideologies, such as Bénabou (2008).

2 The Model

2.1 Set-up

Time is discrete and infinite, with each time period indexed by $s \geq 1$. At each s, the same stage game is played, in which two parties compete for the support of a unit mass of citizens. Time periods are linked via the state variable σ_s, the set of narratives available at the end of period s. This state variable evolves endogenously as parties deploy narratives which voters may or may not adopt.

In the next subsection, we describe the stage game where, for simplicity, we drop the time subscript s since optimal choices depend on time only via the set of narratives contained in the state variable σ_s.

2.2 The stage game

Citizens, policies, and narratives There is a unit mass of citizens comprised of two types, indexed by $K \in \{L, H\}$, where the proportion of type L citizens is denoted by μ. Citizens of type K have a continuous utility function $u^K : \mathbb{R}^m \times \mathbb{R}^l \to \mathbb{R}$ with two arguments: the policy vector $q \in \mathcal{Q} \subset \mathbb{R}^m$ with $m \geq 1$, and a parameter vector $\beta \in \mathcal{B} \subset \mathbb{R}^l$ with $l \geq 1$ that determines the perceived utility effects of policies. Both the set of possible policy vectors $\mathcal{Q}$ and possible parameter vectors $\mathcal{B}$ are assumed to be finite. We denote the utility of a citizen of type K from a specific policy vector q under parameters β by $u^K(q; \beta)$. Policies can differentially affect citizens of different types with the perceived effects depending on the narratives that they adopt.

Narrative demand The parameter vector β determines how policy q affects utility. In standard approaches, this vector is assumed to be disciplined by data which pins down a point estimate of β. Our approach allows for a range of possible parameter values. Beliefs are determined endogenously via the supply and demand of narratives: citizen-specific representations of the causal effects of q.

More formally, a *narrative* ω is a subjective causal model represented as a probability distribution over the set $\mathcal{B}$. Let $P_\omega(\beta)$ be the weight that narrative ω attaches to parameter vector β. For each ω, the weights sum to one across the set of possible parameters, such that $\sum_{\beta \in \mathcal{B}} P_\omega(\beta) = 1$. These weights could be thought of as probabilities over a state space, or as weights assigned to different scenarios based on listening to different accounts of how the world works. We call ω a *conviction narrative* if it assigns all of the weight to a single parameter vector

β so that $P_\omega(\beta) = 1$.[5]

Let

$$U^K(q;\omega) = \sum_{\beta \in \mathcal{B}} P_\omega(\beta)\, u^K(q;\beta) \tag{1}$$

be the utility from narrative ω and policy q.

In equilibrium, citizens choose a narrative to maximize their utility. In any time period s, citizens can choose from the set of available narratives σ_s supplied by politicians, as we explain below. We will focus on symmetric equilibria where citizens of the same type choose the same narrative, which then has to satisfy:

$$\hat{\omega}^K = \arg\max_{\omega \in \sigma_s} \tilde{\pi}^K U^K\left(\tilde{q}_H^K;\omega\right) + \left(1-\tilde{\pi}^K\right) U^K\left(\tilde{q}_L^K;\omega\right) \tag{2}$$

where $\tilde{\pi}^K$ is the conjectured equilibrium probability that party H wins of citizens of type K given conjectured equilibrium policies $\left\{\tilde{q}_L^K, \tilde{q}_H^K\right\}$. Since each individual citizen is of measure zero and does not therefore influence equilibrium outcomes, individuals take these as given when choosing their preferred narrative. An equilibrium requires that these conjectures are correct.[6]

The main non-standard feature of the model is that citizens "choose" their preferred subjective causal model based on a narrative that will influence their policy preferences.[7] Their collective choice then influences the equilibrium policies of parties.

We focus on situations that are common in policy analysis, where an accepted evidence-based correct causal belief, the "truth", is subject to disagreement. This gives license for motivated reasoning to play a role. While we could allow beliefs to be driven *both* by motivated reasoning and accuracy motives, this would only add notational and analytical complexity without fundamentally changing the insights.[8]

[5]We borrow the term "conviction narrative" from Tuckett and Nikolic (2017) and Johnson et al. (2020). Tuckett and Nikolic (2017, p. 502) suggest that "conviction narratives enable actors to draw on the information, beliefs, causal models, and rules of thumb situated in their social context to identify opportunities worth acting on, to simulate the future outcome of the actions by means of which they plan to achieve those opportunities, and to feel sufficiently convinced about the anticipated outcomes to act."

[6]See requirements (3) and (5) below. Henceforth, to keep the notation simple, we no longer refer to the conjectures, since in our equilibrium analysis they need to be correctly anticipated.

[7]As in other models of motivated reasoning, such as Bénabou (2015) or Little et al. (2022), this process of narrative adoption could be thought of as happening subconsciously.

[8]We could, for example, assume that citizens factor in accuracy alongside motivated reasoning when forming beliefs, such as a weighted average between the truth and the utility maximizing belief, as in Little (2019). This would produce very similar results on narrative polarization or narrative consensus, even if *some* weight were placed on the "true" belief. We could also consider

An additional reason to abstract from a desire for accuracy in forming beliefs is the limited feedback onto payoffs from adopting a non-truthful narrative, even if the latter exists. If voters adopt an incorrect belief, and vote for a party whose policy lowers their utility rather than raising it, they still cannot unilaterally swing the election thereby limiting the feedback of their decision onto their payoffs. Beyond this analytical reason, it also agrees with an established body of literature that argues that voters have weak incentives to form correct beliefs about the consequences of complex policy topics (Caplan, 2008). All that matters for our model is that voters are motivated reasoners *at the time of the election*, basing their voting decisions on the beliefs that maximize their prospective utility when they vote. We do not require voters to continue holding these beliefs after the election. Hence, our approach is consistent with voters changing their beliefs as long as at the time of election they hold the narrative that maximizes their prospective utility.

Parties Two office-motivated parties, denoted by $J \in \{L, H\}$, compete for office. Parties L and H are staffed by citizen-candidates of the corresponding income group, who cannot commit to policies ex-ante. After elections, the only incentive compatible policies are those that maximize the utility of the income type that they represent, given the narrative that that income type believes in. It follows that equilibrium party platforms satisfy:

$$\hat{q}_J(\hat{\omega}^J) = \arg\max_q U^J\left(q; \hat{\omega}^J\right), \tag{3}$$

where $\hat{\omega}^J$ is the equilibrium narrative of citizens of group J, as discussed above (see requirement 2). In other words, politicians can only commit to policies which are optimal given their adopted narratives.

Narrative supply While parties cannot commit to policies, they can affect election outcomes through supplying narratives. Let Ω be the finite set of possible narratives from which politicians can choose, which is a superset of the set of narratives σ_s from which citizens choose at time s. Politicians can shape the latter, starting from a predefined initial set of narratives $\sigma_0 \subseteq \Omega$. At the start of each period $s \geq 1$, each party J decides whether to propagate a new narrative $\omega_J \in \Omega \setminus \sigma_{s-1}$. Parties take their decisions to propagate a new narrative simultaneously. Denote party J's decision to propagate a new narrative ω given the set of inherited

a friction such as an idiosyncratic mental costs of adopting "inaccurate" beliefs which would not fundamentally change any insights, except that the results would be driven by the subset of agents for which such mental costs are high while others' beliefs would reflect the "truth".

narratives σ_{s-1} by the indicator variable $I_J^{\omega}(\sigma_{s-1}) \in \{0,1\}$, where 1 means that party J propagates narrative ω.

The set Ω describes the narratives that are in principle acceptable to citizens. We deliberately leave open what constrains this set. One interpretation is that parameters are disciplined by data with the parameter space being bounded by confidence bands around available estimates of a given parameter, or coming from conflicting estimates given by different relevant sources of evidence. Another reason for there being a set of parameters is that the data and assumptions over the data generating process yield only partially identified parameter values, and that a set of parameters is therefore consistent with the evidence (Kline and Tamer, 2023).[9] A non-standard interpretation would be that Ω can also be thought of as comprising the "Overton Window" of what is acceptable in political discourse. Whatever the interpretation, our framework applies to situations where citizens in principle accept a (limited) set of parameter values when forming their beliefs.

If a party deploys a new narrative, then with probability $p \in (0,1)$ it becomes visible and hence part of σ_s for all subsequent periods. We do not assume that narratives disappear from memory in our model, although the logic and main results of our model would be reinforced if the oldest narratives were dropped from the set σ_s whenever a new narrative is added.[10] The probability p can be thought of as representing the likelihood that a narrative is phrased in a way that is convincing or is intelligible to voters. This could depend, for example, on whether a party leader is charismatic or uses persuasive rhetoric in the framing of the narrative. Politicians like Reagan, Thatcher, Clinton and Blair were famous for their abilities to frame their versions of neoliberalism in ways that voters found understandable and appealing.[11]

We assume that the cost of deploying a narrative is δ. This could be an advertising cost or a reduced-form representation of reputation loss for the party that advocates a new narrative. The cost could also represent an internal fight within a party between ideologues and opportunists.

[9] A related idea is that when imposing limited assumptions that do not allow the unique identification of treatment effects, it may still be possible to construct upper and lower bounds on treatment effects (Manski, 1990).

[10] Indeed, our hysteresis result is *harder* to achieve with perfect memory of all past narratives.

[11] If no narrative is supplied at time s, $\sigma_s = \sigma_{s-1}$ with certainty. If one new narrative ω_1 is supplied, the state changes to $\sigma_{s-1} \cup \{\omega_1\}$ with probability p and does not change otherwise. If two narratives are supplied, with probability p^2 the state changes to $\sigma_{s-1} \cup \{\omega_1\} \cup \{\omega_2\}$, with probability $p(1-p)$ it changes to $\sigma_{s-1} \cup \{\omega_1\}$ or $\sigma_{s-1} \cup \{\omega_2\}$, and with probability $(1-p)^2$ it does not change.

Party objectives Parties are myopic and supply narratives to maximize their probability of winning in the current election net of the cost of deploying narratives:

$$\pi_J(q_H(\omega^H), q_L(\omega^L); \omega^H, \omega^L) - \delta \sum_{\omega \in \Omega} I_J^\omega, \tag{4}$$

where the sum of indicator variables has at most one non-zero element since we are restricting parties to supplying at most one narrative in each period. The probability of winning π_J is based on standard probabilistic voting assumptions, described below in equation (5).[12]

There are two key elements to our party model. The first is that parties maximize (4) so care about policies only insofar as they influence their chances of winning. This follows a long-standing tradition in political economy that can be traced back to Downs (1957).[13] Our assumption that politicians are opportunistic is consistent with our assumption that parties cannot commit to a particular policy, choosing the policy that maximizes the payoffs of their core supporters after the election.

The second element is myopia, with parties maximizing their current rather than future payoffs. This makes the model more tractable as well as capturing the intuition that, while the party may be long-lived, individual politicians who control parties have an eye on getting into office and hence focus on short-term considerations. With such myopic behavior, the dynamics come not from forward-looking strategic actions, but from an ever-evolving set of narratives that drive the dynamics of political equilibria.

Given these party objectives, deploying a new narrative has two strategic effects. First, it can influence the narratives that citizens take up, ω^H and ω^L, influencing the perceived utility from any given policy. Second, if citizens buy into a new narrative, parties may change their ex-post optimal policies represented by $q_H(\omega^H), q_L(\omega^L)$. Both of these effects shape the probability of winning through citizens' voting decisions.

Finally, we assume as a tie-breaking rule that if any two strategies give a party the same payoff, they will choose a pure strategy that maximizes the welfare of citizens of their income type. This assumption does not materially impact our

[12]Note that the payoff in (4) is stochastic at the time when the party makes its narrative supply decision since the success of deploying a narrative is probabilistic.

[13]This approach is often contrasted with policy-motivated citizen-candidates (Besley and Coate, 1997; Osborne and Slivinski, 1996). Izzo et al. (2023) develop such a case where parties care about policy and their strategy influences narratives ("ideologies" in their terminology).

results, but disqualifies some Pareto-dominated equilibria in the stage game where parties would get equivalent payoffs as in our main analysis but voters would be worse off.

Voting We use a standard probabilistic voting model as in Persson and Tabellini (2002) with both an idiosyncratic ϵ pro-party H shock that is uniformly distributed on $\left[-\frac{1}{2\phi}, \frac{1}{2\phi}\right]$ and an aggregate pro-party H shock, η, that is uniformly distributed on $\left[-\frac{1}{2\kappa}, \frac{1}{2\kappa}\right]$. Then in equilibrium a voter of type K will vote for party H if

$$U^K(\hat{q}_H; \hat{\omega}^K) + \epsilon + \eta \geq U^K(\hat{q}_L; \hat{\omega}^K).$$

The voting system is majoritarian and the party with half the votes wins the election. So the equilibrium winning probability for party H is given by

$$\hat{\pi}\left(\hat{q}_L, \hat{q}_H; \hat{\omega}^L, \hat{\omega}^H\right) = \frac{1}{2} + \kappa\left[\mu \Delta U^L\left(\hat{q}_H, \hat{q}_L; \hat{\omega}^L\right) + (1-\mu)\,\Delta U^H\left(\hat{q}_H, \hat{q}_L; \hat{\omega}^H\right)\right] \quad (5)$$

where $\Delta U^K\left(\hat{q}_H, \hat{q}_L; \hat{\omega}^K\right) = U^K(\hat{q}_H; \hat{\omega}^K) - U^K(\hat{q}_L; \hat{\omega}^K)$ is the utility difference for a type K citizen from voting for party H rather than party L. We will assume that this has an interior solution (i.e. κ is low enough).

Narratives versus policies While both narratives and policies are vectors over which parties compete for votes, they differ in important ways. For citizens, they differ because they experience only the policy of the winning party. But they do not have to adopt the narrative of the winning party. For parties, the narratives that the citizens adopt constrain their ex-post optimal policy choices. Hence, parties can only influence their electoral prospects by affecting the narratives that citizens adopt.

Timing within the stage game The timing of the stage game is as follows:

1. Each party simultaneously chooses which (if any) new narrative to supply.
2. Citizens choose a narrative from the available set.
3. Citizens vote.
4. The winning party implements policy.
5. Payoffs are realized.

Equilibrium The solution concept that we use is a Markov Perfect Equilibrium where strategies are a function of the state σ_s, and where the stage game is a Subgame Perfect Equilibrium in each period. Thus, for each period s, the equilibrium comprises: a pair of narratives supplied by parties $\{\hat{\omega}_L, \hat{\omega}_H\}$ that maximize current-period party payoffs, given the set of narratives from the previous period

σ_{s-1}; a pair of citizen narratives $\{\hat{\omega}^L, \hat{\omega}^H\}$ adopted by each group that satisfy (2) given the available narratives σ_s; a pair of policy vectors $\{\hat{q}_L, \hat{q}_H\}$ that satisfy (3) given the pair of equilibrium citizen narratives; and citizens' voting choices that anticipate the equilibrium policies and winning probability $\hat{\pi}$ satisfying (5).

3 Analysis

In this section, we analyze some general features of our model, which yields two main insights. First, the presence or absence of *polarizing narratives* drives the strategic actions of parties and the dynamics of equilibrium policies. Second, temporary economic changes can have permanent effects on equilibrium policies. This serves to show how our model differs from existing approaches and provides the background for the application in section 4.

Existence of symmetric equilibria. We start with establishing the existence of symmetric Markov Perfect Equilibria. For this purpose, we define *rationalizability*[14] of a policy vector q, using equation (3) to write the policy that a party chooses as a function of the narrative adopted by citizens in its corresponding income group, i.e. $q_J(\omega^J)$. This yields:

Definition 1. *A policy vector q is **rationalizable for party J in state** σ_s if $\exists\, \omega^J \in \sigma_s : q_J(\omega^J) = q$ and $\omega^J = \arg\max_{\omega' \in \sigma_s} U^J(q; \omega')$.*

In the absence of policy commitment, a policy vector q can only be implemented by party J if there is a narrative ω such that, if income group J believes it, party J would wish to implement q according to (3). Our definition of rationalizability further requires that the same narrative ω is the optimal choice for citizens if they expect policy q to be implemented with certainty.

The following theorem gives sufficient conditions for a *symmetric pure-strategy equilibrium* to exist, where all citizens of the same income group make the same narrative choices in pure strategies:

Theorem 1.

1. *A symmetric pure-strategy equilibrium exists where both parties offer the same policy if there is at least one policy vector q that is rationalizable for both parties.*

[14] This concept of rationalizability differs from the standard conception in game theory that a strategy is rationalizable if it survives iterated deletion of strictly dominated strategies. Our concept of rationalizability instead concerns the relationship between narratives and policy. It is related since, if the first requirement $\exists\, \omega^J \in \sigma_s : q_J(\omega^J) = q$ is not met, then q can never be an optimal choice for party J.

2. *A symmetric pure-strategy equilibrium exists where parties offer different policies* $\hat{q}_L = q_A \neq \hat{q}_H = q_B$ *if there is a narrative* ω^A *that rationalizes* q_A *for party* L *and satisfies* $\omega^A = \arg\max_{\omega' \in \sigma_s} U^L(q_B; \omega')$*, and there is a narrative* ω^B *that rationalizes* q_B *for party* H *and satisfies* $\omega^B = \arg\max_{\omega' \in \sigma_s} U^H(q_A; \omega')$.

For the first statement, note that if the same policy vector is rationalizable for both parties, then coordination on it is always a self-fulfilling prophecy, i.e. if all citizens believe that the policy will be implemented, they will adopt the narratives that rationalize that policy, which leads to it being implemented. The second statement requires that narratives that maximize the citizens' utility of at least two policy vectors exist, which leads to policy divergence in equilibrium. While these conditions are sufficient for the existence of symmetric equilibria in pure strategies, the equilibrium need not be unique.[15]

In the remainder of this section, we will assume that the conditions for the existence of symmetric pure-strategy equilibria are satisfied without explicitly re-stating them each time. The application in section 4 imposes sufficient structure on the policy model to guarantee the existence of such equilibria.

Policy consensus and polarization In a symmetric equilibrium, we can write the narrative choice of a citizen from group J as a function of the state, $\hat{\omega}^J(\sigma_s)$, and a party J's policy choices as a function of citizens' narrative choices (and hence the state) $\hat{q}_J(\hat{\omega}^J(\sigma_s))$. Some states give rise to a *policy consensus*, where both parties converge on the same equilibrium policies:

Definition 2.

(a) σ_s *is a* ***state of policy consensus*** *iff both parties offer the same policies in equilibrium:* $\hat{q}_L(\hat{\omega}^L(\sigma_s)) = \hat{q}_H(\hat{\omega}^H(\sigma_s))$*. Otherwise,* σ_s *is a* ***state of policy polarization****.*

(b) *In a state of policy consensus* σ_s*,* $\omega \in \sigma_s$ *is a* ***consensus narrative*** *iff all citizens adopt it and both parties choose the same policy.*

A state of policy consensus can arise even without a consensus narrative: citizens can adopt different narratives but still have the same policy preferences. Hence, a consensus narrative is sufficient, but not necessary, for the existence of a state of policy consensus.

Starting from a state of policy consensus, new narratives that are successfully deployed could break the consensus and lead to policy polarization. We call these *polarizing narratives*. Whether a narrative is polarizing depends on the state σ_s, since narratives may be polarizing in some states but not in others. More formally:

[15] This is discussed in more detail in Supplemental Appendix B.

Definition 3. *$\omega \in \Omega \setminus \sigma_s$ is a **polarizing narrative** under a state of policy consensus σ_s iff under state $\sigma'_s := \sigma_s \cup \{\omega\}$ the two parties would adopt different equilibrium policies, $\hat{q}_L(\hat{\omega}^L(\sigma'_s)) \neq \hat{q}_H(\hat{\omega}^H(\sigma'_s))$, and these policies would give different levels of utility for at least one income type.*

A polarizing narrative in state σ_s is one that does not yet form part of the state but, if successfully deployed, would lead to policy polarization. In our model, policy polarization can arise in two different ways. The first is conventional, and is a situation where both citizens might adopt the same narrative but with different policies being preferred by the two income groups. This mirrors what happens in standard models of policy divergence where citizens agree on the causal effects of policies but their interests are different. The second is novel and showcases the role of narratives. It occurs when citizens adopt different narratives and this leads to different policy choices. We will explore more precisely how this works in the application in section 4. The final part of our definition of polarizing narratives disqualifies the knife-edge cases of narratives that would lead to different policies with exactly the same utility for all citizens.

Narrative disruption from a consensus state We begin our analysis by considering what happens starting from a state of policy consensus. This is the relevant starting point for our application in section 4, where we consider narrative development starting in the post-war consensus that emerged after WWII. Polarizing narratives play a key role in such cases since, in their absence, no new narratives will ever be deployed. We state this formally in:

Theorem 2. *Starting in a state of policy consensus, there are two possible types of equilibrium paths:*

1. *If no polarizing narrative exists, then no new narratives are ever supplied in equilibrium.*
2. *If a polarizing narrative exists, then for sufficiently low narrative deployment costs δ, at least one new narrative is deployed along any equilibrium path.*

To see this, observe that, in a state of policy consensus, both parties offer the same policies and are therefore indistinguishable to voters. Hence, each party wins with probability one-half. Any narrative that does not lead to policy polarization, either because it keeps the policy consensus or because it generates a new consensus, will lead to the same probability of winning while being costly for parties to deploy. Hence, if there is no polarizing narrative, no party will wish to deploy a narrative.

The second part of the Theorem deals with the complementary case where at least one polarizing narrative does exist. As polarization will lead to an electoral advantage for one party relative to a state of policy consensus, the existence of a polarizing narrative necessitates that one party must find it optimal to deploy it as long as deployment costs δ are sufficiently low. Note that the narrative that does arise in equilibrium need not be a polarizing narrative. Instead it could be a new consensus narrative that preempts polarization. Even then, a necessary condition for a party to wish to deploy a new consensus narrative is that there is a polarizing narrative that *it* preempts. The application in section 4 develops a particular example to illustrate this logic in the context of deregulation policy.

Narrative consensus from a polarized state. Theorem 2 describes equilibrium paths starting from a state of policy consensus. The flipside is to consider what would happen starting from a state of policy polarization. Here, different paths are possible. If σ_s is a state of policy polarization, then one party will have an electoral advantage. The other party's best response then depends on the space of narratives that have not yet been deployed, $\Omega \setminus \sigma_s$. If this set includes a consensus narrative, then the party can close the electoral gap by deploying it. Better yet is a situation where there exists another polarizing narrative that would flip the electoral advantage towards the party that deploys it. Hence, the party that is at a disadvantage under polarization may either try to build a new consensus to close the electoral gap, or to generate its own polarizing narrative that is more favorable. Regardless of its choices, if a consensus narrative exists that can combat polarization, the state of polarization is unstable. How this instability is resolved depends on the features of the policy model that is being studied and we explore this feature in more detail in our application in section 4, for a particular set of assumptions on what the narrative space Ω and the policy q represents.

Policy hysteresis Theorem 2 established that, starting in a state of policy consensus, new narratives can emerge only if at least one narrative can polarize. Whether a narrative can polarize is itself state dependent, opening the door to policy hysteresis where temporary changes can permanently shift policy. To see this, suppose that there is a temporary shock that changes how narratives affect utility, for instance an economic crisis that lowers incomes. This can make some narratives more appealing and some less appealing to different income groups. Importantly, it can create the potential for a polarizing narrative that previously did not exist. As discussed above, one party will have an incentive to deploy it, while the other party will best respond by finding a different polarizing equilibrium

or a new policy consensus. Once these narrative dynamics play out, the equilibrium policy can look very different to what it was before the shock. Importantly, the new narratives can remain citizens' choices even after the shock has subsided that created the narrative dynamic in the first place, preventing a return to the previous policy equilibrium.

How these dynamics play out depends on model details best studied in a particular application. Below, we lay out a specific model to enrich our understanding of how narratives have been evolving after the post-war consensus that came out of WWII. It helps to solve an apparent puzzle that a narrative adopted in times of economic stress remained in place in a more benign economic climate.

4 Application to Neoliberal Narratives

We now explore how our framework can be usefully applied to concrete policy settings. To this end, we develop a specific application to study the emergence of neoliberal narratives, a particular constellation of ideas that took root in the 1980s and displaced a post-war policy consensus. We begin by making precise what *we* mean by "neoliberalism", a term that is often only vaguely defined. We will characterize the conditions for narrative consensus or polarization across groups of citizens and show how the success of the neoliberal narrative depends on the state of the economy, explaining why it became attractive after the turbulence of the 1970s. We then show how the deployment of the neoliberal narrative prompts a new consensus narrative to emerge, which permanently alters the political equilibrium. This allows us to illustrate why hysteresis is possible in a dynamic model of narratives. These arguments are useful in tracing through the dynamics of policy change in a situation where we believe that narrative dynamics played an important role.[16]

4.1 The economic model

Assume that the two types of citizens differ in their private pre-tax incomes α^L and α^H, where $\alpha^L < \alpha^H$. In line with standard classifications in political economy, we refer to citizens of group L as "low income" and of group H as "high income".

[16] In Besley et al. (2026), we explore this claim empirically by looking at policy narratives in speeches made by politicians in the House of Commons in the UK. We find evidence of narrative shifts in line with our model application, suggesting that narrative dynamics may indeed be a feature of dynamic policy change.

Correspondingly, we will refer to the high-income party as the *Right* and the low-income party as the *Left*.

Two policy dimensions affect private incomes: a tax t and a regulation policy x.[17] In the baseline model, for simplicity of exposition we assume that both parties set the same tax t.[18] Throughout, parties can choose regulation policy $x \in [\underline{x}, \overline{x}]$, where higher x corresponds to *de*regulation.[19]

The real income of group K as a function of deregulation is

$$y^K(x; \psi, \lambda) = \alpha^K + \lambda^K(x; \psi, \lambda), \tag{6}$$

where

$$\lambda^K(x; \psi, \lambda) = \begin{cases} \lambda x & \text{for } K = H \\ -\psi\lambda x & \text{for } K = L. \end{cases}$$

The rich always get higher income with deregulation while it has the opposite effect on the income of the poor. This is needed to create a conflict of interest over deregulation policy. The parameter $\lambda \geq 0$ measures the importance of deregulation in shaping incomes, while the parameter $\psi \geq 0$ captures the negative distributive effect on the poor.

Aggregate income is denoted by

$$\begin{aligned} Y(x; \psi, \lambda) &\equiv (1-\mu)\, y^H(x; \psi, \lambda) + \mu y^L(x; \psi, \lambda) \\ &= \alpha + \lambda x(1 - \mu - \psi\mu), \end{aligned}$$

where $\alpha \equiv (1-\mu)\,\alpha^H + \mu\alpha^L$. Aggregate income is increasing in x if $\psi < (1-\mu)/\mu$, i.e. if the negative distributive effect is small enough. We call this inequality the *trickle-down condition*; if it holds, then part of the benefits of deregulation to the rich will trickle down in the shape of higher public good provision to the poor.

Public goods and taxation The government finances a public good G via taxes. Total public goods provision satisfies $G = \gamma t Y(x; \psi, \lambda)$ where $\gamma \geq 0$ captures the efficiency and/or valuation of public goods. If γ is low, then the production of public goods is inefficient, either because taxes cannot be collected without a deadweight loss, or because public goods are not highly valued. The opposite

[17] The economy is modeled similarly to Besley and Persson (2023).

[18] In Supplemental Appendix C, we show that our main results extend to the case where parties can choose a tax $t \in [\underline{t}, \overline{t}]$. Then neoliberalism motivates tax cuts as well as deregulation.

[19] The continuous policy space of x is different and more general than our assumption of a finite policy space in section 2, but if we instead assumed binary $x \in \{\underline{x}, \overline{x}\}$ our results would be very similar.

is true for high γ.

Individuals receive utility from their private post-tax income as well as the provision of public goods G. Utility for an individual from income group K is given by:

$$u^K(t, x; \psi, \lambda, \gamma) = (1 - t)y^K(x; \psi, \lambda) + \gamma t Y(x; \psi, \lambda). \tag{7}$$

The two effects of deregulation Our formulation suggests two channels through which deregulation affects payoffs. The first (direct) effect differs by income, with lower income citizens losing and higher income citizens benefiting from deregulation. The second (indirect) effect of deregulation impacts all citizens in the same way through changing *average* income and hence the level of the public good provided. The latter is the source of trickle down benefits to lower income citizens and captures a key tenet of neoliberal thinking, that gains to the rich can be a benefit to all, in a simple and tractable way.[20]

4.2 The narrative space

Assume that the causal parameters that make up a narrative have binary values, i.e. $\lambda \in \{\underline{\lambda}, \overline{\lambda}\}$, $\psi \in \{\underline{\psi}, \overline{\psi}\}$ and $\gamma \in \{\underline{\gamma}, \overline{\gamma}\}$, where upper-bars denote strictly larger values than lower-bars. We will focus exclusively on conviction narratives, which assign full weight to one particular narrative vector (λ, ψ, γ). Under relatively mild assumptions, this is without loss of generality.[21]

Hence, the narrative space is $\Omega = \{\underline{\lambda}, \overline{\lambda}\} \times \{\underline{\psi}, \overline{\psi}\} \times \{\underline{\gamma}, \overline{\gamma}\}$, giving eight possible conviction narratives. However, to start with, we focus on two of these. The first we call the *post-war consensus narrative*:

$$\omega^c = \{\lambda^c, \psi^c, \gamma^c\} = \{\underline{\lambda}, \overline{\psi}, \overline{\gamma}\}, \tag{8}$$

and the second the *neoliberal narrative:*

$$\omega^n = \{\lambda^n, \psi^n, \gamma^n\} = \{\overline{\lambda}, \underline{\psi}, \underline{\gamma}\}. \tag{9}$$

Comparing these two narratives, observe that (i) $\lambda^n > \lambda^c$ so that the neoliberal narrative proposes that deregulation has a larger impact on incomes than does

[20] It is not, however, the only way of capturing this idea. In Supplemental Appendix C we suggest an alternative trickle-down mechanism based on "individualistic" versus "collectivist" worldviews and show that delivers essentially similar insights.

[21] In particular, if we assume that all the conviction narratives are part of Ω, then for any political equilibrium where citizens of group J adopt mixed narratives, there exists an equilibrium with the same policy outcomes where citizens adopt conviction narratives.

the post-war consensus narrative, (ii) $\psi^n < \psi^c$ so that the post-war narrative proposes that deregulation has a more negative distributive impact on low-income individuals and (iii) $\gamma^n < \gamma^c$ so that the neoliberal narrative views the state as less efficient and public goods as less important than the post-war consensus narrative. Although stylized, we believe that these definitions capture core elements of these views and, in particular, give a precise interpretation of terms that are usually left vague.

We will also work with some specific parametric restrictions in order to home in on what we consider to be the interesting cases to study. The following set of conditions ensures that narrative divergence implies policy divergence, which gives narrative choice bite in our framework:[22]

$$\begin{aligned} t\underline{\gamma}(1-\mu(1+\underline{\psi})) &\geq \underline{\psi}(1-t) \\ &\text{and} \\ (1-t) &\leq t\underline{\gamma}(\mu(1+\overline{\psi})-1). \end{aligned} \tag{10}$$

These conditions imply that the trickle-down condition holds for the neoliberal narrative, such that higher x is believed to increase aggregate income, but not for the post-war consensus narrative.

We also assume that $\overline{\lambda}\underline{\psi} \geq \underline{\lambda}\overline{\psi}$, which implies that the low-income type suffers more from deregulation with the neoliberal narrative than the consensus narrative.

4.3 A stage game with only two narratives

As a starting point for the full analysis, we study citizen choice with $\sigma_s = \{\omega^n, \omega^c\}$, i.e. the choice between the neoliberal narrative and the post-war consensus narrative. In the next subsection, we will turn to the dynamic model and choice between all possible narratives. To study the political equilibrium, we work backwards starting with the policy choice.

Policy Choices At the end of the stage game, each party J implements the policy that maximizes the utility of the income type J they represent, given their adopted narrative ω^J as given by (3). Given condition (10), this implies that $x_J(\omega^c) = \underline{x}$ and $x_J(\omega^n) = \overline{x}$ for $J \in \{L, H\}$. In other words, either party implements deregulation if the citizens from their income class believe in neoliberalism, and implements regulation if they believe in the post-war consensus. Since both

[22] In particular, the condition ensures that any narrative that includes $\underline{\psi}$ ($\overline{\psi}$) implies that deregulation (regulation) is ex-post preferred by both types.

parties offer the same policy for a narrative $\omega \in \{\omega^c, \omega^n\}$, from now onwards, we drop the subscript J and write $x(\omega)$ to denote the policy that follows from narrative ω.

Winning Probabilities Given this, the winning probability of party H, as a function of narrative choices $\{\omega^L, \omega^H\}$, is

$$\pi\left(\omega^L, \omega^H\right) = \frac{1}{2} + \kappa\left[\mu \Delta U^L\left(\omega^L, \omega^H\right) + (1-\mu)\,\Delta U^H\left(\omega^L, \omega^H\right)\right], \tag{11}$$

where with slight abuse of notation, we now write the probability of winning as a function solely of narratives rather than narratives and policies, using the fact that narrative choice implies policy choice from (3).

Narrative Choices To study narrative choices, it is useful to define for each income group $J \in \{L, H\}$, the perceived gain from adopting the neoliberal narrative relative to the post-war consensus:

$$\begin{aligned} E^J\left(x(\omega^L), x(\omega^H)\right) \equiv\ & \pi\left(\omega^L, \omega^H\right)\left[U^J(t, x(\omega^H), \omega^n) - U^J(t, x(\omega^H), \omega^c)\right] \\ & + [1 - \pi\left(\omega^L, \omega^H\right)]\left[U^J(t, x(\omega^L), \omega^n) - U^J(t, x(\omega^L), \omega^c)\right] \\ =\ & \pi\left(\omega^L, \omega^H\right)\left[(1-t)\Delta y^J\left(x\left(\omega^H\right)\right) + t\Delta G\left(x\left(\omega^H\right)\right)\right] \\ & + \left[1 - \pi\left(\omega^L, \omega^H\right)\right]\left[(1-t)\Delta y^J\left(x\left(\omega^L\right)\right) + t\Delta G\left(x\left(\omega^L\right)\right)\right] \end{aligned} \tag{12}$$

which consists of two main components denoted by $\Delta y^J(x)$ and $\Delta G(x)$. The first component is the perceived difference in private income across the two narratives given by:

$$\Delta y^J(x) \equiv y^J(x; \underline{\psi}, \overline{\lambda}) - y^J(x; \overline{\psi}, \underline{\lambda}). \tag{13}$$

The second component is the perceived difference in public goods provision between the neoliberal narrative and the post-war consensus given by:

$$\begin{aligned} \Delta G(x) &\equiv \underline{\gamma} Y(x; \underline{\psi}, \overline{\lambda}) - \overline{\gamma} Y(x; \overline{\psi}, \underline{\lambda}) \\ &= -\alpha\left(\overline{\gamma} - \underline{\gamma}\right) + x\left[\overline{\lambda}\underline{\gamma}\left(1 - \mu\left(1 + \underline{\psi}\right)\right) - \underline{\lambda}\overline{\gamma}\left(1 - \mu\left(1 + \overline{\psi}\right)\right)\right]. \end{aligned} \tag{14}$$

The first term in (14) is negative, representing the more pessimistic view of government efficiency under neoliberalism. The second term is positive given condition (10), due to the increase in aggregate income when the trickle-down condition holds; the neoliberal narrative is more optimistic about the scope for deregulation to increase aggregate output and hence public goods.

If expression (12) is positive for an income group, then members of the group will choose to believe the neoliberal narrative; if it is negative, they will choose the post-war consensus narrative. Since policy depends on the beliefs of *all* citizens, second-order beliefs matter when assessing which narrative to believe in, as they influence the pair of narratives (ω^L, ω^H) which citizens use to evaluate $E^J(x(\omega^L), x(\omega^H))$.

The sign and size of $\Delta G(x)$ depends on mean income, α. The higher is mean income, the bigger is the perceived problem of government waste, and the less attractive is the neoliberal narrative. Put differently, there are fewer resources for government to "waste" in a poorer economy, but also more to gain from deregulation, making neoliberalism the more optimistic option under low α. This parameter therefore plays a key role in determining which narrative equilibrium arises:

Proposition 1. *Assume that only the neoliberal narrative and post-war consensus narrative are available, $\sigma_s = \{\omega^n, \omega^c\}$. Then, equilibrium narrative choices of citizens and policy choices of parties can be characterized through three ranges of α with cutoff values $\underline{\alpha} < \overline{\alpha}$:*

1. *If and only if $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, there is an equilibrium with narrative and policy divergence where the high (low) income citizens believe in the neoliberal narrative (post-war consensus narrative), and the high (low) income party delivers deregulation (regulation) if elected;*
2. *if $\alpha \leq \underline{\alpha}$ there is an equilibrium with full consensus on the neoliberal narrative and deregulation;*
3. *if $\alpha \geq \overline{\alpha}$, there is an equilibrium with full consensus on the post-war consensus narrative and regulation.*

Proposition 1 summarizes how the state of the economy, α, influences the political equilibria. In particular, when α is very high or very low, the equilibrium will be one of narrative convergence across high- and low-income citizens, and policy convergence across parties; only with an intermediate range of α is there scope for narrative and policy divergence. The logic can be understood in four steps.

First, observe that if low-income citizens believe in the neoliberal narrative in equilibrium, then high-income citizens must as well. This is because high-income individuals get the same benefit as low-income individuals from the change in public output when believing in the neoliberal narrative. However, they get a strictly larger private benefit. A similar logic implies that if the high-income

citizens believe in the post-war consensus, then so must the low-income citizens. The only kind of divergent beliefs that are possible are therefore where the high-income citizens adopt the neoliberal narrative while the low-income citizens adopt the post-war consensus narrative.

Second, consider what happens if α is very high. Then, the neoliberal narrative's pessimistic view of government efficiency makes it unattractive even to high-income citizens. Specifically, equation (14) shows that compared to the post-war consensus narrative, the lower government efficiency supposed by the neoliberal narrative leads to a loss in public output by $(\overline{\gamma} - \underline{\gamma})\alpha$. With very high α, this loss is so large that it overshadows any private gains to high-income citizens stemming from the neoliberal narrative. Then *a fortiori,* the low-income citizens also oppose the neoliberal narrative, and instead subscribe to the post-war consensus.

Third, consider the case where α is very low. Now, pessimism about public good efficiency with neoliberal narratives does not have a strong negative effect on utility. In contrast, the potential gains in public output from deregulation via the trickle-down effect are large, captured through the second term in equation (14). For low enough α, this trickle-down effect is sufficient to persuade even the low income individuals to believe in the neoliberal narrative. *A fortiori,* the high income individuals will as well.

Fourth, note that for middling values of α we can have that the public output is seen as roughly equal with both the neoliberal and the post-war consensus narrative. In other words, the two terms from equation (14) roughly cancel each other out, so that $\Delta G(x) \simeq 0$. In that case, only the private gains from deregulation matter in forming a narrative preference, leading high income individuals to support the neoliberal narrative and low-income individuals to support the post-war consensus.

Multiple Equilibria? Proposition 1 showed that a polarized equilibrium exists when $\alpha \in (\underline{\alpha}, \overline{\alpha})$. However, this equilibrium need not be unique. This is due to the two-way dependency between narrative and policy choice. Specifically, if citizens anticipate a deregulated world, that makes it more likely that they will choose to believe the neoliberal narrative. Moreover, with more people believing in the neoliberal narrative, the policy is more likely to be deregulation. This may result in multiple equilibria and we detail the specific parameter ranges for which this is true in Supplemental Appendix B.[23]

[23]This Appendix also makes clear when the polarized equilibrium is unique, which requires conditions on the ordering of E^J functions (see Supplemental Appendix Proposition B.1). Thus, while multiple equilibria are possible, we do get uniqueness under reasonable restrictions that

Extensions Supplemental Appendix C discusses two extensions to the framework. First, we consider what happens when both deregulation x and tax policies t are endogenous. We give conditions under which neoliberalism comes with a tax cut, leading to polarization in both deregulation and tax policies. The main take-away from endogenising tax rates is that, although things get somewhat more complicated, the same broad conclusions still follow as when taxes are fixed. Divergence in equilibrium narratives and policies, which now include both taxation and regulation, occur for intermediate values of α.

Second, we consider an alternative micro-foundation for the trickle-down mechanism. In the baseline model trickle-down runs through the provision of public goods which generates a feedback from the aggregate economy to private payoffs. An alternative way to conceptualize differences between the neoliberal and post-war consensus narrative would be to suppose that these capture differences in individualistic versus collectivist worldviews.[24] The appendix shows that modeling things this way preserves the key insights of the approach.

4.4 Narrative Dynamics

Having solved the stage game for the case where the only narratives to choose from are the post-war consensus narrative ω^c and the neoliberal narrative ω^n, we now examine the full dynamic game. We continue to assume that the narrative space is given by $\Omega = \left\{\underline{\lambda}, \overline{\lambda}\right\} \times \left\{\underline{\psi}, \overline{\psi}\right\} \times \left\{\underline{\gamma}, \overline{\gamma}\right\}$ so that there are eight possible conviction narratives that can be deployed by parties. Among them are ω^c and ω^n that we have already studied. One additional narrative that we have not considered so far turns out to play an important role in the analysis:

$$\omega^3 = \{\overline{\lambda}, \underline{\psi}, \overline{\gamma}\} \tag{15}$$

We refer to ω^3 as the "Third Way" narrative, following a term that has been used to describe the politics of Clinton, Blair, and other politicians from traditionally centre-left parties who adopted a more pro-market stance in the 1990s (Giddens, 1998).[25] This is a form of "left-wing neoliberalism" that combines elements of the neoliberal narrative ω^n and the post-war consensus narrative ω^c. Our formalization

are also intuitively appealing.

[24] This brings this aspect of narratives closer to that suggested in Bénabou (2008).

[25] Interestingly, Gerstle (2022) uses the term "Third Way" as a description of the New Deal narrative that emerged in the 1930s. The term was also used by Wilhelm Röpke, one of the Presidents of the Mont Pelerin Society, to describe an economic order that combines market-based approaches with some role for state intervention (Burgin, 2012).

of it is based on our reading of the shift in policy debates that occurred in the mid 1990s. Specifically, the Third Way narrative ω^3 agrees with neoliberalism on the benefits of deregulation (via a low ψ and high λ), and agrees with the post-war consensus on the efficiency and importance of government spending (via high γ).

To avoid trivial cases where parties never wish to propagate new narratives, we assume that the cost of narrative deployment δ is not too high. In particular, this requires $\delta/p + 1/2 < \max\{\overline{\pi}, 1-\overline{\pi}\}$, where $\overline{\pi}$ is defined in equation (18) below.

The Dynamic Path The following Proposition describes the different paths that are possible in the dynamic model. We assume that in the initial period, only the post-war consensus narrative is available. The logic then mirrors our core analysis in Theorem 2 that described equilibrium paths as a function of whether a polarizing narrative exists. That α affects polarization was shown in Proposition 1 when comparing the neoliberal narrative to the post-war consensus.

To make the analysis less taxonomic, we home in on one particular ordering of the E^J functions that arises if the following condition holds:

$$\begin{aligned} &\overline{\lambda}(\underline{x} + \underline{\psi}\overline{x}) - \underline{\lambda}(\underline{x} + \overline{\psi}\overline{x}) \\ &> \frac{t(\overline{x} - \underline{x})}{1-t}\left(\underline{\gamma}\overline{\lambda}(1 - \mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi}) - 1)\right) \end{aligned} \tag{16}$$

This implies that deregulation impacts income groups sufficiently differently such that the E^H functions intersect 0 at a higher value of α than the E^L functions. We discuss this in greater detail in Supplemental Appendix B, where we also discuss why, even if condition (16) does not hold, our analysis would be similar, although more taxonomic since more cases would need to be considered (see Supplemental Appendix Proposition B.1).

Proposition 2. *Assume that the initial state only contains the post-war consensus narrative. The dynamic equilibrium depends on the state of the economy represented by α:*

1. *If the economy is doing well ($\alpha \geq \overline{\alpha}$), no party ever propagates a new narrative, the state always remains $\sigma_s = \{\omega^c\}$, and there is convergence on regulation across both parties.*
2. *If the economy is doing less well ($\alpha \in (\tilde{\alpha}, \overline{\alpha})$, with cutoff $\tilde{\alpha} < \overline{\alpha}$), then on the equilibrium path one party deploys the neoliberal narrative ω^n with positive probability, the other deploys the Third Way narrative ω^3 with positive probability, and no other narratives are deployed. State $\{\omega^c, \omega^n\}$ is reached*

with positive probability, leading to narrative divergence where only high income citizens believe in the neoliberal narrative and policy divergence where only the high income party supports deregulation. Eventually, one of the two absorbing states $\{\omega^c, \omega^3\}$ *or* $\{\omega^c, \omega^n, \omega^3\}$ *is reached, leading to a narrative consensus on* ω^3*, policy consensus on deregulation, and no other narratives being propagated.*

Two different types of equilibrium paths can be distinguished, depending on the economy. Case 1 parallels our finding from Proposition 1 about the stage game where there was convergence on the post-war consensus. However, that proposition considered only a choice between the neoliberal narrative and the post-war consensus; Proposition 2 shows that this is a global result where the post-war consensus prevails even after allowing for the set of all possible narratives to be propagated by parties. With high enough α, no polarizing narrative exists so that, by Theorem 2, no narrative will then ever be deployed. A consensus on regulation policy is thus maintained.

Case 2 is more interesting and corresponds to a situation where the economy is doing less well as represented by lower α. Now the Proposition says that one party will propagate the neoliberal narrative, and the other the Third Way narrative. Eventually, the state will converge to a new consensus where the Third Way narrative is successfully deployed and all citizens believe in it, with deregulation being the consensus policy of both parties. However, along the dynamic path to this outcome, there is potential for periods in which the neoliberal narrative creates polarization in narratives and policies.

Discussion To explain the logic of Proposition 2, recall that in the initial state σ_0 where only the post-war consensus narrative, ω^c, is available, then it (trivially) is a consensus narrative in which all voters believe in the merits of regulation and both parties converge on $\underline{x}$. Both parties are indistinguishable in the eyes of voters and neither party has an electoral advantage, so each wins with probability one-half.

The only way for a party to gain an electoral advantage is to deploy a polarizing narrative, whose deployment would lead to policy differentiation across parties (see Theorem 2). Starting with the post-war consensus where ω^c is adopted by all citizens, only the neoliberal narrative ω^n can be polarizing for $\alpha > \tilde{\alpha}$. We show in the proof of the Proposition that the remaining narratives cannot be polarizing either because they are too unattractive for any citizen to believe in them or

because they are too attractive and would convince everyone.[26]

A key insight from the model is how the multi-dimensionality of the parameter space plays a role in narrative dynamics. Specifically, the neoliberal narrative can create polarization because it disagrees with the post-war consensus on two dimensions. Disagreement with ω^c on the trickle-down effect through a different assessment of ψ implies policy disagreement. Furthermore, disagreement on government efficiency through a different γ creates an additional cleavage between income groups, increasing the potential for polarization.

It is not inevitable that neoliberalism is polarizing and therefore strategically attractive for a party. Proposition 2 stresses that this depends on the state of the economy α. With high α, neoliberal pessimism in government efficiency is too costly for *either* income group to find it attractive relative to the post-war consensus. Specifically, there is a loss of $-(\overline{\gamma} - \underline{\gamma})\alpha$ from adopting the neoliberal narrative, as can be seen in equation (14). Thus, a strong economy — high α — makes neoliberalism unattractive. Proposition 2 also shows that, starting from a high α, the neoliberal narrative can only break the post-war consensus if there is a downturn in the economy, i.e. a decrease in α. This corresponds to an established view that neoliberal narratives gained traction during the economic turbulence of the 1970s, an idea captured succinctly in the preface of Friedman (1982, p. xiii-xiv):

> "Only a crisis - actual or perceived - produces real change. When that crisis occurs, the actions that are taken depend on the ideas that are lying around. That, I believe, is our basic function: to develop alternatives to existing policies, to keep them alive and available until the politically impossible becomes the politically inevitable."

Narratives can exist in the minds of intellectuals, but they only become part of the *political process* when the conditions are right. The neoliberal narrative was clearly known before economic turbulence hit in the 1970s but our model suggests that it was not going to gain traction even if a party deployed it in the benign economic circumstances that prevailed.[27] In contrast, after a downturn that leads

[26] See the Appendix for the proof to Proposition 2. It outlines some subtleties in the reasoning. We show that some narratives will not be employed because they achieve the same electoral outcomes while being less preferred by citizens. One other narrative is polarizing for a lower range of $\alpha < \tilde{\alpha}$ which we do not consider here; however, it is strictly dominated by ω^n from the citizens' perspective and therefore does not change the long-run outcomes either.

[27] Of course, we do not allow economic shocks to be a function of the policies in place and doubtless some proponents of neoliberalism blamed the turbulence of the 1970s on the policies that prevailed. However, the oil shocks of the 1970s were plausibly exogenous.

to $\alpha \in (\tilde{\alpha}, \overline{\alpha})$, the neoliberal narrative ω^n is attractive to high-income but not low-income citizens. It follows that one party will find it optimal to deploy it.

After the neoliberal narrative has been successfully deployed, leading to policy polarization, the party that deployed the narrative gains an electoral advantage. Since elections are zero-sum games, this puts the opposing party at a disadvantage. To negate this disadvantage, the optimal strategy is to bring forward a new narrative that restores the political fortune of the party that loses from neoliberalism. This is done by deploying the Third Way narrative.

What is special about the Third Way narrative is that it overcomes the neoliberal polarization and creates a new consensus by combining the most optimistic elements from the other competing narratives. It has neoliberal optimism about the value of deregulation as embodied in high λ and low ψ. It has post-war consensus optimism in its views about the effectiveness of government in the form of high γ. It overcomes the Achilles heel of politicians like Reagan and Thatcher that was their pessimism about the role of the state. After the Third Way narrative is successfully deployed, as it was by Clinton, Obama and Blair, it creates a new consensus on deregulation. With both parties converging onto the same policy, voters are again indifferent between parties, and no party benefits from an electoral advantage over the post-war consensus even though the underlying regulation policies are different.

The results illustrate an interesting feature of dynamic political competition with narratives. Starting from a consensus, the neoliberal narrative is deployed along the equilibrium path even though, from the perspective of citizens, it is dominated by the Third Way narrative. Propagating neoliberalism is attractive for a party precisely because it is polarizing and the Third Way is electorally useful only when it is used to counter the neoliberal narrative in order to create a new consensus. Thus, neoliberalism can be thought of as a disruptive narrative to break a consensus.[28]

For the same reason that the Third Way narrative can combat existing neoliberal polarization, it can also preempt it. Whether or not it pays off for a party to try to preemptively propagate the Third Way narrative depends on the probability of success of narratives, captured through the parameter p, as well as the costs of propagation, captured through the parameter δ. How these parameters feed into the dynamic equilibrium depends, however, on which party benefits from the polarized neoliberal equilibrium.

[28]Narrative competition therefore has some features that are familiar from models of creative destruction (Aghion and Howitt, 1992).

When will Neoliberalism be propagated by right-wing parties? Proposition 2 does not say *which* party will propagate the neoliberal narrative. Given that parties are assumed to care about winning this is not a matter of party ideology. Moreover, the policy choice that any party makes simply follows what is optimal for the group that they represent. In reality, we know that parties of the political right, the Republicans in the US and the Tories in the UK, were the vanguard of neoliberalism. We now explore when our model predicts that this will be the case.

The key condition that is needed for it to be the high income party to propagate the neoliberal narrative is:

$$\frac{\overline{\lambda}}{\underline{\lambda}} > \frac{\mu}{1-\mu}\frac{(1-t)\overline{\psi}+t\overline{\gamma}(\mu(1+\overline{\psi})-1)}{(1-t)+t\underline{\gamma}(1-\mu(1+\underline{\psi}))} \tag{17}$$

The right party receives additional votes from high-income voters who adopt the neoliberal narrative and hence prefer deregulation, but fewer votes from low-income voters who adhere to the post-war consensus narrative and support regulation. The gain from increased high-income votes exceeds the loss from low-income voters when (17) holds. Whenever $\overline{\lambda}$ is large enough relative to $\underline{\lambda}$, party H gains from narrative polarization.

In this polarized equilibrium where high income individuals adopt the neoliberal narrative, party H's winning probability is:

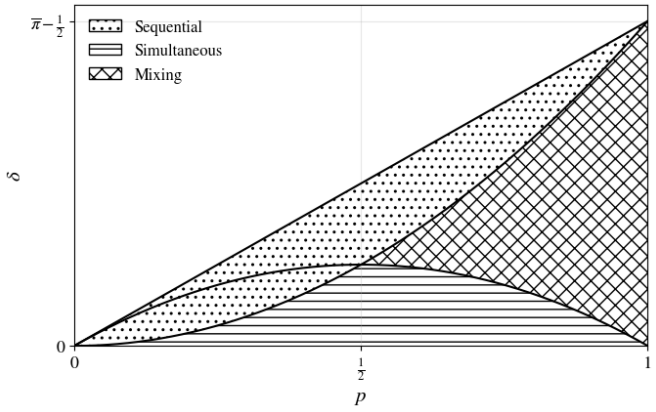
$$\begin{aligned}\overline{\pi} \equiv \frac{1}{2} + \kappa(\overline{x}-\underline{x})[&(1-\mu)\overline{\lambda}\left((1-t)+t\underline{\gamma}(1-\mu(1+\underline{\psi}))\right)\\ &-\mu\underline{\lambda}\left((1-t)\overline{\psi}+t\overline{\gamma}(\mu(1+\overline{\psi})-1)\right)],\end{aligned} \tag{18}$$

which is greater than a half when (17) holds. Correspondingly, the left's winning probability is then $1-\overline{\pi}$. The left benefits from countering the neoliberal narrative to increase its winning chances back to one-half. However, whether it benefits from doing so preemptively depends on the probability that narrative deployment succeeds p and the cost of deploying narratives δ. This is made precise in:

Proposition 3. *Suppose that condition (17) holds and that $\alpha \in (\tilde{\alpha}, \overline{\alpha})$. Then there are three cases:*

1. *If $\delta \in (p^2(\overline{\pi}-0.5), p(\overline{\pi}-0.5))$, on the unique equilibrium path, the right party propagates narrative ω^n in state $\{\omega^c\}$, the left party propagates narrative ω^3 in state $\{\omega^c, \omega^n\}$, and no narratives are propagated in state $\{\omega^c, \omega^n, \omega^3\}$.*
2. *If $\delta < \min[p^2(\overline{\pi}-0.5), (1-p)p(\overline{\pi}-0.5)]$, in state $\{\omega^c\}$ the right party*

Figure 1: Three types of equilibria when $\alpha \in (\tilde{\alpha}, \overline{\alpha})$



Note: This figure shows the three ranges of p and δ that correspond to the three types of equilibria from Proposition 3. "Sequential" refers to case 1 from Proposition 3, "Simultaneous" to case 2, and "Mixing" to case 3.

propagates ω^n and the left propagates narrative ω^3. On the equilibrium path, the state either changes to $\{\omega^c, \omega^n\}$ where the left propagates ω^3 and then to $\{\omega^c, \omega^n, \omega^3\}$ where no narratives are propagated; or it changes directly to $\{\omega^c, \omega^n, \omega^3\}$ or $\{\omega^c, \omega^3\}$, where no narratives are propagated.

3. *If $(1-p)p(\overline{\pi}-0.5) < \delta < p^2(\overline{\pi}-0.5)$, in state $\{\omega^c\}$ the right party mixes between propagating ω^n and not propagating any new narrative, and the left party mixes between propagating ω^3 and not propagating any narrative. Possible equilibrium paths are the same as in case 2 above.*

This proposition gives the conditions required for the left to propagate the Third Way narrative preemptively along the dynamic path, *before* the right deploys the neoliberal narrative. The key intuition is that the payoff from propagating ω^3 differs from that of propagating ω^n because ω^n generates an electoral benefit if it is deployed successfully while the Third Way narrative is not (which occurs with probability $p(1-p)$). In contrast, ω^3 is only electorally valuable if neoliberalism *and* the Third Way are both deployed successfully (which occurs with probability p^2). Whenever $p \neq 1/2$ these two payoffs are not the same. Thus, the left will preemptively propagate the Third Way narrative only if it is not too costly to do so (low δ), or if narrative deployment is highly likely to be successful (high p). If δ is sufficiently high or p is sufficiently low, the Third Way narrative emerges only *after* neoliberalism has been deployed successfully by the right party. Figure 1 summarizes these insights, showing the ranges of δ and p that correspond to the three cases in Proposition 3.

Hysteresis Our model also illustrates how temporary economic shocks can have permanent effects on policy via changing the quality of existing narratives. This offers an insight into why there was no reversion to the post-war consensus even after the arrival of benign economic conditions from the mid 1990s onwards, often called the Great Moderation.

To see this more clearly, suppose that the average (pre-regulation) wages differ across periods, given by α_s. The state at time s is then given by the set of narratives and the economy, $\{\sigma_s, \alpha_s\}$, and politicians and voters observe the state of the economy at the start of each stage game. At the end of the stage game, the equilibrium policy $\hat{x}_s$ is implemented by the winning party.

The next proposition shows how this leads to hysteresis when starting in a post-war consensus equilibrium in period 0:

Proposition 4. *Suppose that in the initial state $\alpha_0 \geq \overline{\alpha}$ and $\sigma_0 = \{\omega^c\}$, such that both parties initially agree on regulation ($\hat{x}_0 = \underline{x}$).*

1. *If $\alpha_s > \overline{\alpha}$ for all time periods s, no new narratives are offered ($\sigma_s = \{\omega^c\}$) and deregulation is never undertaken ($\hat{x}_s = \underline{x}$).*
2. *If there is a temporary shock such that $\alpha_s \in (\tilde{\alpha}, \overline{\alpha})$ for $k \geq 1$ consecutive periods and $\alpha_s > \overline{\alpha}$ for the remaining periods, then, for low enough δ, the Third Way narrative is deployed with positive probability, leading to a permanent policy consensus on deregulation ($\hat{x}_s = \overline{x}$) if it is deployed.*

The first part of the proposition follows from Case 1 in Proposition 2. In the absence of a temporary economic shock, the post-war consensus remains in place because no polarizing narrative exists. The second part of the Proposition describes what happens after a temporary shock to α_s, which makes it optimal for one party to deploy the neoliberal narrative. Then, in line with the logic of Case 2 in Proposition 2, the Third Way will be deployed with positive probability before the economy recovers, after which it will lead to a new consensus on deregulation. Importantly, *even after the economy recovers*, citizens will prefer the Third Way narrative over the post-war consensus, and both parties will continue to support deregulation.

This observation reveals a special feature of a model where political dynamics are driven by shifts in narratives since, after a narrative is propagated and becomes available to voters, it remains an option. If narratives live on in citizens' memories after they are deployed, they remain part of the currency of future political campaigns. And politicians will only be successful in employing new narratives if they are able to dislodge existing ones. Motivated reasoning therefore becomes

a source of policy friction. This observation perhaps explains the frustration of those who supported the post-war consensus in the UK and US and have been harsh critics of politicians like Clinton, Obama and Blair for their perpetuation of the Third Way which they (correctly) observe preserves many elements of neoliberalism. With motivated reasoning, there is no way to persuade the voters to return to the post-war model starting from a world in which the Third Way narrative has been successfully deployed.

5 Concluding Comments

This paper has laid out a framework for studying the role of narratives as subjective causal models in a canonical model of political competition of the kind that is frequently used in the political economy literature: two parties from the left and right are competing for power and class differences are the key political cleavage.

We have explored static and dynamic aspects of a world in which citizens can choose which narratives to incorporate into their beliefs based on self-serving motivated reasoning. This can result in narrative polarization where preferences for divergent policies are driven not just by interests but by differing views about how the economy works. Our dynamic framework shows that temporary economic shocks can give rise to narrative dynamics that permanently alter policies.

We have argued that this kind of framework can usefully be applied to study the conditions for neoliberal thinking and policy making. While initiated by an economic shock in the 1970s, the persistence of neoliberalism has been a source of puzzlement. As De Long (2023) has put it: "The smart thing to say is that ... neoliberalism is, or ought to be, dead, but zombie-like, it still walks the earth." But the detachment of narrative-based beliefs from evidence and the use of motivated reasoning can allow such persistence to emerge and to be sustained. Our approach can also explain why a consensus formed around a particular version of neoliberalism which has been referred to as the "Third Way" since Clinton and Blair.

Although these are complex issues, we think that there is value in debating them using a modeling framework that can bring a degree of specificity and precision to what are often quite vague concepts. In spite of the inevitably stylized nature of the approach that we have suggested with only two groups, two parties and two policy dimensions, there is richness in the analysis, especially when dynamics are considered.

Our framework focused exclusively on the neoliberal narratives in domestic

politics. It is silent on the international diffusion of its ideas under the so-called "Washington Consensus" which propagated these ideas globally through the World Bank and IMF. A large literature investigates how pro-market policies spread across countries (see Buera et al., 2011), but the mechanisms by which this happens remains to be analyzed formally. Insofar as the domestic success of neoliberalism in some countries such as the US and UK is a precondition for its international spread, our framework offers a first step towards understanding what became a globally deployed narrative.

We have homed in on a specific application where we believe that narrative dynamics were an important driver of policy change. Expanding the scope of this will require bespoke models that capture salient features of specific policy areas. For instance, future applications can investigate new narratives that have emerged in more recent decades and have been supplanting the Third Way. In line with the logic of our model, shocks such as the Great Recession or COVID-19 pandemic could have generated new polarizing narratives along policy dimensions not discussed in our application.

We have modeled narratives as being isomorphic to subjective causal models without getting into their cognitive underpinnings. However, exploring these is an important topic for future work on narratives. Arguably what makes a narrative influential is its capacity to make specific causal features of policy salient (Bordalo et al., 2020, 2012) or creating something that is memorable when voters are casting their ballots. Kendall and Charles (2022) study a micro-founded model with experimental evidence and it could be fruitful to connect this to policy case studies which could provide a tighter connection to the literature on cognition and bounded rationality in policy choice (Bendor, 2010). This should improve our understanding of the frictions that allow them to prevail in the face of evidence, what makes deployment of narratives successful and what sets the limits on the set of available narratives.

There is also an important empirical counterpart to this theoretical research program that uses recent advances in Large Language Models (LLMs) to analyze text data. Our model of neoliberal narratives suggests three key parameters that comprise the narrative that can be explored in political manifestos or parliamentary speeches. Besley et al. (2026) use LLMs to examine speeches in the UK House of Commons to study one of these three elements, narratives on the size of government, showcasing some patterns in line with the model: a post-war consensus on Big Government that persists until Margaret Thatcher's Conservative Party decisively broke for Small Government, which initiated a period of polarization

on the size of government that persisted for several decades. With the current pace of advances in LLMs, it becomes ever easier to analyze political narratives holistically, and in a more structured way.[29]

References

Aghion, Philippe, and Peter Howitt. 1992. A Model of Growth Through Creative Destruction. *Econometrica* 60(2): 323–351.

Aina, Chiara. 2023. Tailored stories. Available at `https://chiaraaina.github.io/files/TailoredStories_Aina.pdf`.

Alesina, Alberto, Armando Miano, and Stefanie Stantcheva. 2023. Immigration and redistribution. *The Review of Economic Studies* 90(1): 1–39.

Alesina, Alberto, and Stefanie Stantcheva. 2020. Diversity, Immigration, and Redistribution. *AEA Papers and Proceedings* 110: 329–34.

Ash, Elliott, Germain Gauthier, and Philine Widmer. 2024. Relatio: Text semantics capture political and economic narratives. *Political Analysis* 32(1): 115–132.

Asirvatham, Hemanth, Elliott Mokski, and Andrei Shleifer. 2026. GPT as a Measurement Tool. NBER Working Paper 34834.

Bénabou, Roland. 2008. Ideology. *Journal of the European Economic Association* 6(2-3): 321–352.

Bénabou, Roland. 2015. The economics of motivated beliefs. *Revue d'économie politique* (5): 665–685.

Bénabou, Roland, Armin Falk, and Jean Tirole. 2018. Narratives, imperatives, and moral reasoning. NBER Working Paper 24798.

Bénabou, Roland, and Jean Tirole. 2016. Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives* 30(3): 141–164.

Bendor, Jonathan. 2010. *Bounded rationality and politics*. Univ of California Press.

Besley, Timothy, Adam Brzezinski, and Jake Fazzio. 2026. Economic policy narratives: a taxonomy and application. *LSE Public Policy Review* 4(1).

Besley, Timothy, and Stephen Coate. 1997. An economic model of representative democracy. *The quarterly journal of economics* 112(1): 85–114.

[29]For instance, Asirvatham et al. (2026) offer a new tool that provides a more standardized way to leverage large language models that could be used to study political narratives based on parliamentary speeches.

Besley, Timothy, and Torsten Persson. 2023. Origins of New Political Cleavages: Education, Incomes, and Values. CEPR: Discussion Paper No.18667.

Bonomi, Giampaolo, Nicola Gennaioli, and Guido Tabellini. 2021. Identity, beliefs, and political conflict. *The Quarterly Journal of Economics* 136(4): 2371–2411.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. Salience theory of choice under risk. *The Quarterly journal of economics* 127(3): 1243–1285.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2020. Memory, attention, and choice. *The Quarterly journal of economics* 135(3): 1399–1442.

Breznau, Nate, Eike Mark Rinke, Alexander Wuttke, Hung H. V. Nguyen, et al.. 2022. Observing many researchers using the same data and hypothesis reveals a hidden universe of uncertainty. *Proceedings of the National Academy of Sciences* 119(44): e2203150119.

Brzezinski, Adam. 2023. Narratives of Blame. *Oxford DPhil Chapter* 6–90.

Buera, Francisco J., Alexander Monge-Naranjo, and Giorgio E. Primiceri. 2011. Learning the Wealth of Nations. *Econometrica* 79(1): 1–45.

Burgin, Angus. 2012. *The great persuasion: Reinventing free markets since the depression*. Harvard University Press.

Bursztyn, Leonardo, Georgy Egorov, Ingar Haaland, Aakaash Rao, and Christopher Roth. 2023a. Justifying dissent. *The Quarterly Journal of Economics* 138(3): 1403–1451.

Bursztyn, Leonardo, Aakaash Rao, Christopher Roth, and David Yanagizawa-Drott. 2023b. Opinions as facts. *The Review of Economic Studies* 90(4): 1832–1864.

Caplan, Bryan. 2008. *The myth of the rational voter: why democracies choose bad policies-new edition*. Princeton University Press.

Chinoy, Sahil, Nathan Nunn, Sandra Sequeira, and Stefanie Stantcheva. 2023. Zero-sum thinking and the roots of US political divides. NBER Working Paper 31688.

Daunton, Martin. 2022. The Oil Shock and Neo-Liberalism. Available at `https://www.gresham.ac.uk/watch-now/oil-shock`.

De Long, Brad. 2023. Ask a Neoliberal: An Interview with J. Bradford DeLong. *Dissent* Fall 2023. Available at `https://www.dissentmagazine.org/article/ask-a-neoliberal-an-interview-with-j-bradford-delong/`.

Downs, Anthony. 1957. An economic theory of democracy. *Harper and Row* 28.

Durante, Ruben, Paolo Pinotti, and Andrea Tesei. 2019. The political legacy of entertainment TV. *American Economic Review* 109(7): 2497–2530.

Eliaz, Kfir, Simone Galperti, and Ran Spiegler. 2022. False narratives and political

mobilization. *arXiv preprint arXiv:2206.12621* .

Eliaz, Kfir, and Ran Spiegler. 2020. A Model of Competing Narratives. *American Economic Review* 110(12): 3786–3816.

Friedman, Milton. 1951. Neo-Liberalism and its Prospects. *Farmand*, 17 February 1951, pp. 89-93; accessed at `https://miltonfriedman.hoover.org/internal/media/dispatcher/214957/full`.

Friedman, Milton. 1982. *Capitalism and Freedom*. University of Chicago Press, 2nd edition.

Gentzkow, Matthew, Jesse M Shapiro, and Matt Taddy. 2019. Measuring group differences in high-dimensional choices: method and application to congressional speech. *Econometrica* 87(4): 1307–1340.

Gerstle, Gary. 2022. *The rise and fall of the neoliberal order: America and the world in the free market era*. Oxford University Press.

Giddens, Anthony. 1998. *The Third Way: The Renewal of Social Democracy*. Polity Press.

Glaeser, Edward L. 2005. The political economy of hatred. *The Quarterly Journal of Economics* 120(1): 45–86.

Harrison, Brian. 1999. The rise, fall and rise of political consensus in Britain since 1940. *History* 84(274): 301–324.

Hoff, Karla, Jyotsna Jalan, and Sattwik Santra. 2021. *Participatory theater empowers women: Evidence from India*. World Bank.

Izzo, Federica, Gregory J Martin, and Steven Callander. 2023. Ideological Competition. *American Journal of Political Science* 67(3): 687–700.

Johnson, Samuel GB, Avri Bilovich, and David Tuckett. 2020. Conviction Narrative Theory: A Theory of Choice under Radical Uncertainty. *Behavioral and Brain Sciences* 1–47.

Kamenica, Emir. 2019. Bayesian persuasion and information design. *Annual Review of Economics* 11(1): 249–272.

Kamenica, Emir, and Matthew Gentzkow. 2011. Bayesian persuasion. *American Economic Review* 101(6): 2590–2615.

Kearney, Melissa S, and Phillip B Levine. 2015. Media influences on social outcomes: The impact of MTV's 16 and pregnant on teen childbearing. *American Economic Review* 105(12): 3597–3632.

Kendall, Chad W, and Constantin Charles. 2022. Causal narratives. NBER Working Paper 30346.

Keynes, John Maynard. 1936. The General Theory of Employment, Interest and Money.

Kline, Brendan, and Elie Tamer. 2023. Recent developments in partial identification. *Annual Review of Economics* 15(1): 125–150.

Kunda, Ziva. 1990. The case for motivated reasoning. *Psychological bulletin* 108(3): 480.

La Ferrara, Eliana, Alberto Chong, and Suzanne Duryea. 2012. Soap operas and fertility: Evidence from Brazil. *American Economic Journal: Applied Economics* 4(4): 1–31.

Levy, Gilat, Ronny Razin, and Alwyn Young. 2022. Misspecified Politics and the Recurrence of Populism. *American Economic Review* 112(3): 928–62.

Little, Andrew T. 2019. The distortion of related beliefs. *American Journal of Political Science* 63(3): 675–689.

Little, Andrew T., Keith E. Schnakenberg, and Ian R. Turner. 2022. Motivated reasoning and democratic accountability. *American Political Science Review* 116(2): 751–767.

Manski, Charles F. 1990. Nonparametric bounds on treatment effects. *The American Economic Review* 80(2): 319–323.

Montiel Olea, José Luis, Pietro Ortoleva, Mallesh M Pai, and Andrea Prat. 2022. Competing models. *The Quarterly Journal of Economics* 137(4): 2419–2457.

Montiel Olea, José Luis, and Andrea Prat. 2025. Competing Ideologies: Fit, Simplicity, and Fear. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5188291`.

Murphy, Kevin M, and Andrei Shleifer. 2004. Persuasion in politics. *American Economic Review* 94(2): 435–439.

Ortoleva, Pietro. 2022. Alternatives to bayesian updating. *Annual Review of Economics* 16.

Osborne, Martin J, and Al Slivinski. 1996. A model of political competition with citizen-candidates. *The Quarterly Journal of Economics* 111(1): 65–96.

Persson, Torsten, and Guido Tabellini. 2002. *Political economics: explaining economic policy*. MIT press.

Reagan, Ronald. 1981. Address to the Nation on the Economy. Accessed from the Archives of the Ronald Reagan Presidential Library and Musuem.

Schwartzstein, Joshua, and Adi Sunderam. 2021. Using models to persuade. *American Economic Review* 111(1): 276–323.

Shayo, Moses. 2009. A model of social identity with an application to political economy: Nation, class, and redistribution. *American Political science review* 103(2): 147–174.

Shiller, Robert J. 2017. Narrative economics. *American economic review* 107(4):

967–1004.

Slobodian, Quinn. 2018. *Globalists: The end of empire and the birth of neoliberalism.* Harvard University Press.

Stantcheva, Stefanie. 2024. Perceptions, Mindsets and Beliefs Shaping Policy views. 2024 Economica-Coase Lecture.

Stedman Jones, Daniel. 2014. *Masters of the Universe: Hayek, Friedman, and the birth of neoliberal politics-updated edition.* Princeton University Press.

Tuckett, David, and Milena Nikolic. 2017. The role of conviction and narrative in decision-making under radical uncertainty. *Theory & psychology* 27(4): 501–523.

A Proofs

A.1 Theorem 1

For the first point, denote by ω^L and ω^H the narratives that rationalize q for party L and H, respectively. Suppose that in equilibrium $\hat{q}_L = \hat{q}_H = q$. Since q is implemented with certainty, equilibrium citizen narratives must maximize $U^J(q;\omega)$ for each income group J. By the definition of rationalizability, ω^J maximizes this utility and makes the policy vector q the choice of party J, as required.

For the second statement, suppose that in equilibrium $\hat{q}_L = q_A$ and $\hat{q}_H = q_B$. Then, equilibrium narratives of type J citizens maximize the expected utility $\hat{\pi}_L U^J(q_A;\omega) + (1-\hat{\pi}_L)U^J(q_B;\omega)$. As narrative ω^A maximizes $U^L(q_A;\omega)$ (by the definition of rationalization) and maximizes $U^L(q_B;\omega)$ (by assumption stated in the Theorem), it also maximizes the expected utility for citizens of type L. Equivalently, ω^B maximizes the expected utility for citizens of type H. Finally, by the definition of rationalization, ω^A makes q_A the choice for party L and ω^B makes q_B the policy choice for party H, as required.

A.2 Theorem 2

For the first statement, note that in a state of policy consensus the probability of winning is 1/2 for each party. If there does not exist a polarizing narrative, then any other narrative will lead to the same probability of winning, and hence no narrative can be deployed. It follows that neither party deploying a narrative must be an equilibrium, and exactly one party deploying a narrative with positive probability cannot be an equilibrium. For uniqueness of the equilibrium, note that

the condition does not say that there exists no polarizing narrative in a state where both parties deploy a narrative. Suppose that such a polarizing equilibrium does exist when both parties deploy a new narrative with strictly positive probability, and suppose that it leads to one party having a winning probability of more than 1/2, which is the only relevant case. But then, the other party would be strictly better off had it not deployed the narrative. It follows that in equilibrium neither party deploys a narrative.

For the second statement, if a polarizing narrative exists, then after the deployment one party would have a higher probability from winning. It follows that for δ low enough, there cannot be an equilibrium where neither party supplies a narrative. Towards contradiction, suppose that neither party supplies a narrative in equilibrium. Then the party that stands to gain from polarization will, by supplying a narrative, increase with probability p its winning probability, at a cost δ. For sufficiently small δ, the party that stands to gain from supplying the narrative would do so. Hence, in any equilibrium, at least one party will supply a narrative with strictly positive probability when δ is sufficiently low.

A.3 Proposition 1

The proof proceeds in three steps. We first argue that the only possible equilibria are $(\omega^L, \omega^H) \in \{(\omega^c, \omega^c), (\omega^c, \omega^n), (\omega^n, \omega^n)\}$. Next, we show that the E^J functions can be ordered, so that $E^L(x(\omega^c), x(\omega^c)) < E^L(x(\omega^c), x(\omega^n)) < E^H(x(\omega^c), x(\omega^n)) < E^H(x(\omega^n), x(\omega^n))$. Finally, we define the thresholds $\underline{\alpha}$ and $\overline{\alpha}$, and characterize the equilibria for different ranges of α.

Set of equilibria To see that there cannot be an equilibrium where only low-income citizens believe in neoliberalism, recall that the payoff from believing in the neoliberal narrative for group J is given by equation (12). Since $\overline{\lambda}\underline{\psi} \geq \underline{\lambda}\overline{\psi}$, $\Delta y^L \leq 0 < \Delta y^H$, and hence for any narrative ω^L, ω^H, the payoff from believing in the neoliberal narrative is higher for high-income relative to low-income citizens. Hence, in equilibrium, it is impossible that low-income citizens adopt the neoliberal narrative while high-income citizens adopt the post-war consensus narrative. It follows that the only three possible narrative equilibria are $(\hat{\omega}^L, \hat{\omega}^H) \in \{(\omega^c, \omega^c), (\omega^c, \omega^n), (\omega^n, \omega^n)\}$.

Ordering of E^J functions Since the payoff from adopting the neoliberal narrative is strictly higher for high-income citizens, we know that $E^H(x(\omega^L), x(\omega^H)) > E^L(x(\omega^L), x(\omega^H))$ for any (ω^L, ω^H). For the remaining ordering, we can now show

that:

$$\begin{aligned} & E^H(x(\omega^n), x(\omega^n)) > E^H(x(\omega^c), x(\omega^c)) \\ \iff & (1-t)\underbrace{(\Delta y^H(\overline{x}) - \Delta y^H(\underline{x}))}_{\text{positive}} + t\underbrace{(\Delta G(\overline{x}) - \Delta G(\underline{x}))}_{\text{positive}} > 0 \end{aligned}$$

We can also show that:

$$\begin{aligned} & E^L(x(\omega^n), x(\omega^n)) > E^L(x(\omega^c), x(\omega^c)) \\ \iff & \overline{\lambda}\underbrace{(t\underline{\gamma}(1-\mu(1+\underline{\psi})) - (1-t)\underline{\psi})}_{\text{positive (implied by condition 10)}} > \underbrace{-\underline{\lambda}(t\overline{\gamma}(\mu(1+\overline{\psi}) - 1) + (1-t)\overline{\psi})}_{\text{negative (implied by condition 10)}} \end{aligned}$$

Since $E^J(x(\omega^c), x(\omega^n)) = \pi(\omega^c, \omega^n)E^J(x(\omega^n), x(\omega^n)) + (1-\pi(\omega^c, \omega^n))E^J(x(\omega^c), x(\omega^c))$, for any $\pi(\omega^c, \omega^n) \in (0,1)$ we have the ordering $E^J(x(\omega^n), x(\omega^n)) > E^J(x(\omega^c), x(\omega^n)) > E^J(x(\omega^c), x(\omega^c))$.

From all of the above, it follows that $E^L(x(\omega^c), x(\omega^c)) < E^L(x(\omega^c), x(\omega^n)) < E^H(x(\omega^c), x(\omega^n)) < E^H(x(\omega^n), x(\omega^n))$, as required.

Characterizing equilibria We now define $\underline{\alpha}$ as the unique α that satisfies $E^L(x(\omega^c), x(\omega^n)) = 0$, and $\overline{\alpha}$ as the unique α that satisfies $E^H(x(\omega^c), x(\omega^n)) = 0$. From the previous arguments, we know that $E^L(x(\omega^c), x(\omega^n)) < E^H(x(\omega^c), x(\omega^n))$ for any given α. We also know that E^J is linearly decreasing in α (see equation (12)). Hence, we know that the cutoffs satisfy $\underline{\alpha} < \overline{\alpha}$. Finally, we know that $\underline{\alpha} > \alpha^L_{cc}$, where the latter is defined as the α that satisfies $E^L(x(\omega^c), x(\omega^c)) = 0$, which by definition satisfies

$$\alpha^L_{cc} t \frac{\overline{\gamma} - \underline{\gamma}}{\underline{x}} = \overline{\lambda}\left(t\underline{\gamma}(1-\mu(1+\underline{\psi})) - (1-t)\underline{\psi}\right) + \underline{\lambda}(t\overline{\gamma}(\mu(1+\overline{\psi}) - 1) + (1-t)\overline{\psi}),$$

which is positive given condition (10). It follows that $\underline{\alpha} > 0$.

To characterize the equilibria, consider the polarized narrative combination (ω^c, ω^n). Note first that in the range $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, $E^H(x(\omega^c), x(\omega^n)) \geq 0$ and $E^L(x(\omega^c), x(\omega^n)) \leq 0$. Hence, for that range, it follows that, given narratives are assumed to be (ω^c, ω^n), low-income citizens prefer ω^c and high-income citizens prefer ω^n. Therefore, (ω^c, ω^n) is an equilibrium. If instead $\alpha < \underline{\alpha}$, $E^L(x(\omega^c), x(\omega^n)) > 0$, and hence low-income citizens would prefer ω^n, and (ω^c, ω^n) does not constitute an equilibrium. Similarly, if instead $\alpha > \overline{\alpha}$, $E^H(x(\omega^c), x(\omega^n)) < 0$, and hence high-income citizens would prefer ω^c, and (ω^c, ω^n) does not constitute an equilibrium. Hence, (ω^c, ω^n) constitute equilibrium narratives if and only if $\alpha \in [\underline{\alpha}, \overline{\alpha}]$,

as stated in the Proposition.

Now, consider the equilibrium (ω^n, ω^n). Following the logic above, this constitutes an equilibrium if $E^L(x(\omega^n), x(\omega^n)) \geq 0$ and $E^H(x(\omega^n), x(\omega^n)) \geq 0$. Given $E^L(x(\omega^n), x(\omega^n)) < E^H(x(\omega^n), x(\omega^n))$ and since E^J is downwards sloping in α, this is satisfied for all $\alpha \leq \alpha^L_{nn}$, where the latter is defined as the α that satisfies $E^L(x(\omega^n), x(\omega^n)) = 0$. Now, because $E^L(x(\omega^c), x(\omega^n)) < E^L(x(\omega^n), x(\omega^n))$, it follows that $\underline{\alpha} < \alpha^L_{nn}$. Since (ω^n, ω^n) is an equilibrium for all $\alpha \leq \alpha^L_{nn}$, *a fortiori*, it is also an equilibrium for all $\alpha \leq \underline{\alpha}$, as stated in the Proposition.

Equivalent reasoning to the previous paragraph establishes that (ω^c, ω^c) constitutes an equilibrium for all $\alpha \geq \overline{\alpha}$, as stated in the Proposition.

A.4 Proposition 2

The proof proceeds in three steps.

Step 1: Ruling out narratives We can show that several narratives will never be propagated in equilibrium, given that the narrative ω^c is always available. First, $\{\underline{\lambda}, \overline{\psi}, \underline{\gamma}\}$: because $\underline{\gamma}$ gives strictly lower utility than $\overline{\gamma}$ and all other parameters are the same as in ω^c, this narrative will never be adopted by any citizens in equilibrium in any state that contains ω^c, and hence will never be deployed.

Second, $\{\overline{\lambda}, \overline{\psi}, \overline{\gamma}\}$: by condition (10), given $\overline{\psi}$, utility is strictly decreasing in λ. Hence, ω^c is preferred by all citizens to this narrative, and hence this narrative can never be deployed.

Third, $\{\overline{\lambda}, \overline{\psi}, \underline{\gamma}\}$: this is dominated by ω^c for both reasons explained above, and hence will never be deployed.

Finally, $\{\underline{\lambda}, \underline{\psi}, \overline{\gamma}\}$: given condition (10), we know that if citizens support this narrative, their party will implement deregulation. Moreover, having ruled out the three narratives above, we know that the only narrative with $\overline{\psi}$ that can be ever heard is ω^c. The narrative we consider here will always beat ω^c in pairwise comparison, since it matches it but with lower ψ, which always gives higher utility to citizens. Hence, propagating it can never create polarization, but, depending on what other narratives are in σ_s, either not change the policy equilibrium (if no citizens believe ω^c) or create a new consensus on deregulation (if some or all believe ω^c). In the former case, there are no electoral benefits, and hence it will not be deployed. In the latter case, to create policy convergence the party would instead deploy narrative ω^3 as it gives strictly higher utility to all citizens; if ω^3 is already available to citizens, then it is always preferred by all citizens (given utility is increasing in λ conditional on $\underline{\psi}$, from condition (10)) and hence deploying

$\{\underline{\lambda}, \underline{\psi}, \overline{\gamma}\}$ cannot be optimal.

Step 2: Polarizing Narratives Having ruled out four narratives, we are left with ω^n, ω^3 and the narrative $\tilde{\omega}^n \equiv \{\underline{\lambda}, \underline{\psi}, \underline{\gamma}\}$. Out of these, ω^3 is never polarizing as it creates a consensus on deregulation. In turn, ω^n and $\tilde{\omega}^n$ will never be employed to build a new consensus, since for that purpose, ω^3 is strictly better. Hence, ω^n and $\tilde{\omega}^n$ will be employed only if they are polarizing.

Following the same logic as Proposition 1, we can give conditions under which $\tilde{\omega}^n$ is polarizing when compared to ω^c. To do this, we define $\tilde{E}^J$ analogously to E^J (see equation 12), with the exception that we compute the utility difference between $\tilde{\omega}^n$ and ω^c, instead of ω^n and ω^c. Analogously to before, we define $\tilde{\Delta}y^J$ and $\tilde{\Delta}G$. We have that $\tilde{\Delta}y^H(x) = 0$, $\tilde{\Delta}y^L(x) = \underline{\lambda}(\overline{\psi} - \underline{\psi})x > 0$, and $\tilde{\Delta}G(x) = (\underline{\gamma} - \overline{\gamma})\alpha + x\underline{\lambda}(\underline{\gamma}(1 - \mu(1 + \underline{\psi})) + \overline{\gamma}(\mu(1 + \overline{\psi}) - 1))$.

Note that $\tilde{\Delta}y^L(\overline{x}) > \tilde{\Delta}y^L(\underline{x}) > 0 = \tilde{\Delta}y^H(x)$ and $\tilde{\Delta}G(\overline{x}) > \tilde{\Delta}G(\underline{x})$. Hence, for any given equilibrium $\hat{\omega}^L, \hat{\omega}^H$, $\tilde{E}^L(\hat{\omega}^L, \hat{\omega}^H) > \tilde{E}^H(\hat{\omega}^L, \hat{\omega}^H)$. This means that there are three possible combinations of ω^L, ω^H in equilibrium, as $\omega^L = \omega^c, \omega^H = \tilde{\omega}^n$ cannot occur. It also follows that, for each J, $\tilde{E}^J(x(\tilde{\omega}^n), x(\tilde{\omega}^n)) > \tilde{E}^J(x(\tilde{\omega}^n), x(\omega^c)) > \tilde{E}^J(x(\omega^c), x(\omega^c))$. Putting these together yields the ordering $\tilde{E}^L(x(\tilde{\omega}^n), x(\tilde{\omega}^n)) > \tilde{E}^L(x(\tilde{\omega}^n), x(\omega^c)) > \tilde{E}^H(x(\tilde{\omega}^n), x(\omega^c)) > \tilde{E}^H(x(\omega^c), x(\omega^c))$.

Now, note that comparing $E^L(\overline{x}, \overline{x})$ and $\tilde{E}^L(\overline{x}, \overline{x})$, we have

$$E^L(\overline{x}, \overline{x}) - \tilde{E}^L(\overline{x}, \overline{x}) = (\overline{\lambda} - \underline{\lambda})\overline{x}\left(t\underline{\gamma}(1 - \mu(1 + \underline{\psi})) - (1 - t)\underline{\psi}\right) > 0,$$

where the inequality follows from condition (10).

Hence, it follows that $E^L(\overline{x}, \overline{x}) > \tilde{E}^L(\overline{x}, \overline{x})$, and a fortiori, it follows that $E^L(\overline{x}, \overline{x}) > \tilde{E}^L(\overline{x}, \underline{x})$.

From Proposition 1, we know that ω^n is polarizing in the region $\alpha \in [\underline{\alpha}, \overline{\alpha}]$, where the cut-offs are defined in the proof to the proposition. By equivalent reasoning, $\tilde{\omega}^n$ is polarizing in the region $\alpha \in [\underline{\upsilon}, \overline{\upsilon}]$, where $\underline{\upsilon}$ is defined as the α that satisfies $\tilde{E}^H(\overline{x}, \underline{x}) = 0$ and $\overline{\upsilon}$ as the one that satisfies $\tilde{E}^L(\overline{x}, \underline{x}) = 0$.

Now, define $\tilde{\alpha}^L_{nn}$ as the α that satisfies $\tilde{E}^L(\overline{x}, \overline{x}) = 0$, and α^L_{nn} as the one that satisfies $E^L(\overline{x}, \overline{x}) = 0$. We know that $\overline{\upsilon} < \tilde{\alpha}^L_{nn} < \alpha^L_{nn} < \overline{\alpha}$, where the final inequality follows from condition (16). Hence, in the range between $\tilde{\alpha} \equiv \max\{\underline{\alpha}, \overline{\upsilon}\}$ and $\overline{\alpha}$, only ω^n is polarizing.

Step 3: Equilibria From the above, it follows that in the range $\alpha > \overline{\alpha}$, there exists no polarizing narrative. This leads to the first result of the proposition.

In the range $\alpha \in [\tilde{\alpha}, \overline{\alpha}]$, ω^n is the only polarizing equilibrium, and only high-

Table 1: Payoffs in state $\sigma_s = \{\omega^c\}$ and $\alpha \in [\tilde{\alpha}, \overline{\alpha}]$

	$I_2^{\omega^3}(\omega^c) = 0$	$I_2^{\omega^3}(\omega^c) = 1$
$I_1^{\omega^n}(\omega^c) = 0$	$1/2$ $1/2$	$1/2 - \delta$ $1/2$
$I_1^{\omega^n}(\omega^c) = 1$	$1/2 - p(\Pi - 1/2)$ $1/2 + p(\Pi - 1/2) - \delta$	$1/2 - p(1-p)(\Pi - 1/2) - \delta$ $1/2 + p(1-p)(\Pi - 1/2) - \delta$

income citizens adopt it (see Proposition 1). Each party now has three strategies: promoting no narrative, deploying ω^n, or deploying ω^3. If no narrative is deployed, we have convergence on ω^c, and Right's winning probability is $\pi(\omega^c, \omega^c) = 1/2$. If only ω^n is successfully deployed, we have $\omega^H = \omega^n$ and $\omega^L = \omega^c$, and Right's winning probability is $\pi(\omega^c, \omega^n) \equiv \overline{\pi}$ (see equation 18). This can be above or below $\frac{1}{2}$. If ω^3 is successfully deployed, there is convergence on deregulation, and Right's winning probability is $\pi(\omega^3, \omega^3) = 1/2$.

Recall that if a narrative is propagated, it is successfully deployed with probability p, and incurs a cost δ. Depending on whether $\overline{\pi} \lessgtr 1/2$, it will be the right or the left party that potentially benefits from ω^n. Let us denote $\Pi = \max\{\overline{\pi}, 1-\overline{\pi}\}$, and the party that benefits (loses) from ω^n by party 1 (2).

We start by analyzing optimal strategies in state $\sigma_s = \{\omega^c\}$. Note now for party 1, ω^3 is strictly dominated by a mix of ω^n and deploying no narrative, while for party 2, ω^n is strictly dominated by a mix of ω^3 and deploying no narrative. Hence, party 1 chooses between deploying the neoliberal narratives and deploying none, $I_1^{\omega^n}(\omega^c) \in \{0, 1\}$, while party 2 chooses between deploying the Third Way narrative and deploying none, $I_2^{\omega^3}(\omega^c) \in \{0, 1\}$. The payoffs are summarized in the payoff matrix in Table 1.

Note that in any equilibrium, party 1 deploys ω^n with positive probability: Since $\delta < p(\Pi - 1/2)$, no party deploying a narrative cannot be an equilibrium, since then party 1 would deviate to deploying ω^n. Further, it cannot be an equilibrium that party 1 does not deploy a narrative while party 2 does, since then party 2 would deviate to deploy none. Hence, in any equilibrium, party 1 deploys ω^n with positive probability. Party 2 may or may not deploy ω^3 in state $\{\omega^c\}$, as shown in the proof to the next proposition where the same logic applies when swapping the next proof's parties H and L for the current proof's parties 1 and 2, respectively. In state $\{\omega^c, \omega^n\}$, Party 2's payoff from deploying ω^3 relative to not is $p\frac{1}{2} + (1-p)(1-\Pi) - \delta - (1-\Pi) = p(\Pi - 1/2) - \delta$, which is positive, and hence $I_2^{\omega^3}(\omega^c, \omega^n) = 1$.

Hence, in state $\sigma_s = \{\omega^c\}$, ω^n is deployed with positive probability while ω^3

may or may not be deployed with positive probability; in state $\sigma_s = \{\omega^c, \omega^n\}$, ω^3, is deployed; and in state $\sigma_s = \{\omega^c, \omega^3\}$ or $\sigma_s = \{\omega^c, \omega^n, \omega^3\}$ no further narratives are deployed. It follows that the two possible equilibrium paths entail either ω^n becoming available before ω^3, or vice versa.

A.5 Proposition 3

First, to see why the right party would benefit from ω^n under condition (17), note that this condition is equivalent to $\overline{\pi} > 1/2$ (see equation 18).

Second, note that Table 1 from the proof of the previous proposition gives the relevant payoffs, where $\overline{\pi} = \Pi$ given that condition (17) holds.

This leads to the following three cases: (1)$I_H^{\omega^n}(\omega^c) = 1, I_L^{\omega^3}(\omega^c) = 0$ is the unique equilibrium if and only if $\delta > p^2(\overline{\pi} - 1/2)$; (2) $I_H^{\omega^n}(\omega^c) = 1, I_L^{\omega^3}(\omega^c) = 1$ is the unique equilibrium if and only if $\delta < \min\{p^2(\overline{\pi} - 1/2), p(1-p)(\overline{\pi} - 1/2)\}$; (3) there exists a unique mixed strategy equilibrium where both parties mix with strategies $q_L \in (0,1)$, $q_H \in (0,1)$ if and only if $p(1-p)(\overline{\pi}-1/2) < \delta < p^2(\overline{\pi}-1/2)$, where this interval is only nonempty if $p > 1/2$. Further, as explained in the proof to the previous proposition, we have that $I_H^{\omega^n}(\omega^c, \omega^3) = 0$ and $I_L^{\omega^3}(\omega^c, \omega^n) = 1$.

A.6 Proposition 4

The first statement of the proposition directly follows from the Case 1 of Proposition 2, which covers the case when $\alpha > \overline{\alpha}$.

For the second statement, note that whenever $\alpha \in (\tilde{\alpha}, \overline{\alpha})$, it follows from Case 2 of Proposition 3 that party L will deploy ω^3 with certainty if and only if $\delta < \min[p^2(\overline{\pi}-0.5), (1-p)p(\overline{\pi}-0.5)]$, which in any single period will succeed with probability p. Hence, for a shock that lasts k periods, the probability of successful deployment of ω^3 becomes $1-(1-p)^k$. Once it is deployed, then all citizens buy into it as it dominates both ω^n and ω^c (see Proof to Proposition 2), leading either party to choose $\hat{x}_s = \overline{x}$.

Supplemental Appendix

for

A Dynamic Model of Political Narratives and the Rise of Neoliberalism

Timothy Besley
LSE

Adam Brzezinski
LSE

B Multiple equilibria

This section discusses more fully the conditions which give rise to the ordering of $E^J(x(\omega^L), x(\omega^H))$ functions assumed in section 4.4. It also gives a full account of the multiple and unique equilibria for that particular ordering.

Recall that in equilibrium, $(\omega^L, \omega^H) \in \{(\omega^c, \omega^c), (\omega^c, \omega^n), (\omega^n, \omega^n)\}$; hence, there are three $E^J(x(\omega^L), x(\omega^H))$ functions for each $J \in \{L, H\}$. Under certain conditions, $E^J(x(\omega^L), x(\omega^H))$ satisfies the following monotonicity criterion, which makes characterization of multiple equilibria easy: all three $E^L(x(\omega^L), x(\omega^H))$ become negative at lower values of α than any of the three $E^H(x(\omega^L), x(\omega^H))$ functions. Intuitively, this implies that higher income individuals have a higher payoff from believing in the neoliberal narrative than lower income individuals, no matter what they believe other citizens believe. The following Lemma gives the condition under which E^J satisfies this type of monotonicity.

Lemma B.1. *E^J satisfies monotonicity if* (16) *holds.*

Proof. From the proof to Proposition 1, we know that the following ordering is satisfied: $E^L(x(\omega^c), x(\omega^c)) < E^L(x(\omega^c), x(\omega^n)) < E^H(x(\omega^c), x(\omega^n)) < E^H(x(\omega^n), x(\omega^n))$. Hence, for monotonicity, we only require that $E^H(x(\omega^c), x(\omega^c))$ becomes negative for higher α than $E^L(x(\omega^n), x(\omega^n))$. In particular, we have that

$$
\begin{aligned}
&E^H(x(\omega^c), x(\omega^c)) \\
&= (1-t)(\overline{\lambda} - \underline{\lambda})\underline{x} - t(\overline{\gamma} - \underline{\gamma})\alpha + t\underline{x}(\underline{\gamma}\overline{\lambda}(1 - \mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi}) - 1)); \\
&E^L(x(\omega^n), x(\omega^n)) \\
&= -(1-t)(\underline{\psi}\overline{\lambda} - \overline{\psi}\underline{\lambda})\overline{x} - t(\overline{\gamma} - \underline{\gamma})\alpha + t\overline{x}(\underline{\gamma}\overline{\lambda}(1 - \mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi}) - 1)).
\end{aligned}
$$

We define now $\underline{\zeta}$ as the α that satisfies $E^L(x(\omega^n), x(\omega^n)) = 0$ and $\overline{\zeta}$ as the one that satisfies $E^H(x(\omega^c), x(\omega^c)) = 0$. We then have that:

$$
\begin{aligned}
\underline{\zeta}(\overline{\gamma} - \underline{\gamma})/\overline{x} &= -\frac{1-t}{t}(\underline{\psi}\overline{\lambda} - \overline{\psi}\underline{\lambda}) + \underline{\gamma}\overline{\lambda}(1 - \mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi}) - 1) \\
\overline{\zeta}(\overline{\gamma} - \underline{\gamma})/\underline{x} &= \frac{1-t}{t}(\overline{\lambda} - \underline{\lambda}) + \underline{\gamma}\overline{\lambda}(1 - \mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi}) - 1)
\end{aligned}
$$

Based on this, we find that:

$$
\overline{\zeta} > \underline{\zeta} \iff
$$

$$
\frac{1-t}{t}(\overline{\lambda}(\underline{x} + \underline{\psi}\overline{x}) - \underline{\lambda}(\underline{x} + \overline{\psi}\overline{x})) > (\overline{x} - \underline{x})\left(\underline{\gamma}\overline{\lambda}(1 - \mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi}) - 1)\right)
$$

which is condition (16). □

Intuitively, the condition is satisfied whenever $\overline{x}$ is not too large relative to $\underline{x}$, as $E^J(x(\omega^n), x(\omega^n))$ is increasing in $\overline{x}$, and $E^J(x(\omega^c), x(\omega^c))$ is increasing in $\underline{x}$.

Given that condition (16) from Lemma B.1 is satisfied, the following proposition characterizes the ranges under which there are unique and multiple equilibria:

Proposition B.1. *Suppose that condition (16) is satisfied. Then, there are five ranges of α with cut-off values $\underline{\alpha} < \underline{\zeta} < \overline{\zeta} < \overline{\alpha}$ that determine equilibrium beliefs $(\hat{\omega}^L, \hat{\omega}^H)$:*

1. *if $\alpha < \underline{\alpha}$, there is a unique equilibrium with $(\hat{\omega}^L, \hat{\omega}^H) = (\omega^n, \omega^n)$.*
2. *if $\alpha \in [\underline{\alpha}, \underline{\zeta}]$, there are two pure strategy equilibria with $(\hat{\omega}^L, \hat{\omega}^H) \in \{(\omega^n, \omega^n), (\omega^c, \omega^n)\}$.*
3. *if $\alpha \in (\underline{\zeta}, \overline{\zeta})$, there is a unique equilibrium with $(\hat{\omega}^L, \hat{\omega}^H) = (\omega^c, \omega^n)$.*
4. *if $\alpha \in [\overline{\zeta}, \overline{\alpha}]$, there are two pure strategy equilibria with $(\hat{\omega}^L, \hat{\omega}^H) \in \{(\omega^c, \omega^n), (\omega^c, \omega^c)\}$.*
5. *if $\alpha > \overline{\alpha}$, there is a unique equilibrium with $(\hat{\omega}^L, \hat{\omega}^H) = (\omega^c, \omega^c)$.*

Proof. We define $\underline{\alpha}$ and $\overline{\alpha}$ as in the proof to Proposition 1, and $\underline{\zeta}$ and $\overline{\zeta}$ as in the proof to the previous lemma. Given the condition of Lemma B.1 is satisfied and building on our analysis from the proof to Proposition 1, we know that $0 < \underline{\alpha} < \underline{\zeta} < \overline{\zeta} < \overline{\alpha}$.

Note first that for $\alpha < \underline{\alpha}$ (the first range in the Proposition), all E^H functions are positive, and hence it must be that $\hat{\omega}^H = \omega^n$, and hence the two equilibria possibilities are (ω^c, ω^n) and (ω^n, ω^n). By definition of $\underline{\alpha}$, for $\alpha < \underline{\alpha}$, $E^L(x(\omega^c), x(\omega^n))$ is positive, and a fortiori, $E^L(x(\omega^n), x(\omega^n))$ is positive. Hence, the unique equilibrium is (ω^n, ω^n). Equivalent reasoning shows that for $\alpha > \overline{\alpha}$, the unique equilibrium is (ω^c, ω^c).

For $\alpha \in (\underline{\zeta}, \overline{\zeta})$, by definition of $\underline{\zeta}$ and $\overline{\zeta}$, all E^H functions are positive and all E^L functions are negative. Hence, the unique equilibrium is (ω^c, ω^n).

For $\alpha \in [\underline{\alpha}, \underline{\zeta}]$, all E^H functions are positive and hence $\hat{\omega}^H = \omega^n$. The two possible equilibria are then (ω^c, ω^n) and (ω^n, ω^n). In this range of α, $E^L(x(\omega^c), x(\omega^n)) < 0$ and $E^L(x(\omega^n), x(\omega^n)) > 0$, and hence (ω^c, ω^n) and (ω^n, ω^n) are equilibria. Equivalent reasoning shows that for $\alpha \in [\overline{\zeta}, \overline{\alpha}]$, (ω^c, ω^n) and (ω^c, ω^c) are equilibria. □

Proposition B.1 makes it explicit that there are multiple equilibria in two ranges: $\alpha \in [\underline{\alpha}, \underline{\zeta}]$ or $\alpha \in [\overline{\zeta}, \overline{\alpha}]$. In each case, the results are driven by higher-order beliefs about what narratives other people of an income group believe in. As an example, take the lower range in which multiple equilibria arise, $\alpha \in [\underline{\alpha}, \underline{\zeta}]$. In this lower range of α, high-income types always believe in the neoliberal narrative, while lower-income types can believe in either narrative. In particular, for

lower-income types, it pays off to believe in the neoliberal narrative if and only if they think that others of their own income type do as well. If other low-income individuals believe in the neoliberal narrative, they know that the equilibrium policy will be deregulation, in which case it pays off for them to adopt the neoliberal narrative. In contrast, if other low-income individuals believe in the post-war consensus narrative, then regulation will occur whenever the left party wins. For this range of α, it would be then optimal for low-income individuals to adopt the post-war consensus narrative.

If condition (16) does not hold, there still exist ranges of α with unique and multiple equilibria, but the details of these will differ. For instance, there will not exist a range of α with a unique polarized equilibrium, but there will exist ranges of α where either of the three pure-strategy equilibria will arise. Note, however, that the results from Proposition 1 in the main text hold regardless of condition (16).

C Additional Results

Tax Polarization

In section 4 we assumed that tax rates are fixed with $\underline{t} = \overline{t} = t$ in order to focus on the implications of narratives for regulation. However, debates around competing narratives often have implications for taxation with neoliberalism frequently being associated with a plea for lower taxes. *Prima facie*, this makes sense in our framework given that low γ is part of the neoliberal narrative. A complete treatment would require exploring all possible policy combinations including policy divergence on taxes even when there is a consensus on narratives. To avoid being overly taxonomic, we home in on perhaps the most interesting case where those who subscribe to neoliberalism believe in the merits of low taxes and deregulation, with the opposite being true for those who subscribe to the post-war consensus.

To bring this into the model, we allow parties to choose both taxes $t \in [\underline{t}, \overline{t}]$ and deregulation $x \in [\underline{x}, \overline{x}]$. While endogenous taxation does add complications, the thrust of the results discussed in Proposition 1 remain when we allow neoliberal narratives to support tax cuts. Specifically, with high α, there is no scope for neoliberalism, while with intermediate α, neoliberalism creates policy polarization over two dimensions, taxes and regulation.

For this purpose, we assume that condition (10) now holds for $t \in [\underline{t}, \overline{t}]$, associating neoliberalism (the post-war consensus) with a preference for deregulation (regulation). The assumption needed to extend this to tax preferences is:

$$\begin{aligned} \overline{\gamma} \;&\geq\; \frac{\alpha^H + \underline{\lambda}\overline{x}}{\alpha - \underline{\lambda}\overline{x}(\mu(1+\overline{\psi}) - 1)} \\ &\text{and} \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (C.1) \\ \underline{\gamma} \;&\leq\; \frac{\alpha^L - \underline{\psi}\overline{\lambda}\overline{x}}{\alpha + \overline{\lambda}\overline{x}(1 - \mu(1+\underline{\psi}))} \end{aligned}$$

This makes intuitive sense as we need $\overline{\gamma}$ ($\underline{\gamma}$) to be high (low) enough for the high (low) income party to prefer high (low) taxes with the post-war consensus (neoliberalism). Hence, if (C.1), then $\hat{t}_L(\omega^n) = \hat{t}_H(\omega^n) = \underline{t}$ ($\hat{t}_L(\omega^c) = \hat{t}_H(\omega^c) = \overline{t}$).

Thus, when (10) and (C.1) both hold, we can focus on situations where neoliberalism and the post-war consensus are associated with specific policy outcomes.

Compared to the case of fixed taxes, endogenizing taxes leads to two complications. The first one is that the additional assumption (C.1) is a function

of α, which was not the case for assumption (10). Hence, changing α not only changes equilibrium narratives and policies, as it did in Proposition 1, but also potentially invalidates the assumptions necessary for the analysis. We will thus have to impose that assumption C.1 holds for the relevant range of α considered in the proposition, which we did not need to do for Proposition 1 since the relevant assumptions there did not depend on α.

The second complication arises because neoliberalism now leads to a lower tax rate. This means that the perceived gain from believing in the neoliberal narrative captured by E^J is now additionally influenced by whether taxes are higher or lower. Since taxes are lower with neoliberalism, it means that the neoliberal trickle-down benefit through public goods is now lower than it would be if taxes were fixed. This effect diminishes the perceived benefit from public goods provision that comes from believing in the neoliberal narrative, influencing the ordering of the E^J functions. In particular, for lower-income citizens this means that the E^L function may now fail to satisfy the assumption that the payoffs of believing in the neoliberal narrative turns positive for higher α when others believe in it, an assumption that was automatically satisfied with exogenous taxes. The following Lemma provides a condition under which this is satisfied and the ordering of E^J functions is preserved.

Lemma C.1. *The ordering of E^J functions is preserved relative to the case of shared taxes if the following condition holds:*

$$\begin{aligned}&(\overline{x}-\underline{x})\left(\underline{\gamma}\overline{\lambda}(1-\mu(1+\underline{\psi}))+\overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi})-1)\right)\\&>\left(\overline{x}(1-\underline{t})/\underline{t}-\underline{x}(1-\overline{t})/\overline{t}\right)\left(\overline{\lambda}\underline{\psi}-\underline{\lambda}\overline{\psi}\right)\end{aligned}\tag{C.2}$$

Proof. E^J functions are now a function of taxes and regulation and given by:

$$\begin{aligned}&E^J\left(x(\omega^L),x(\omega^H),t(\omega^L),t(\omega^H)\right)\\&=\pi\left(\omega^L,\omega^H\right)\left[(1-t_H(\omega^H))\Delta y^J\left(\hat{x}_H\left(\omega^H\right)\right)+t_H(\omega^H)\Delta G\left(\hat{x}_H\left(\omega^H\right)\right)\right]\\&+\left[1-\pi\left(\omega^L,\omega^H\right)\right]\left[(1-t_L(\omega^L))\Delta y^J\left(\hat{x}_L\left(\omega^L\right)\right)+t_L(\omega^L)\Delta G\left(\hat{x}_L\left(\omega^L\right)\right)\right]\end{aligned}$$

Note that they are still downwards sloping in α through ΔG.

We start with the observation that the payoff in believing in the neoliberal narrative is still higher for high-income than low-income people, for any ω^L, ω^H: as before, since $\overline{\lambda}\underline{\psi} > \underline{\lambda}\overline{\psi}$ we have that $\Delta y^L < 0 < \Delta y^H$. Hence, we have that $E^H(x(\omega^L), x(\omega^H), t(\omega^L), t(\omega^H)) > E^L(x(\omega^L), x(\omega^H), t(\omega^L), t(\omega^H))$.

For the remaining ordering, we first show that $E^H(x(\omega^n), x(\omega^n))$ becomes 0 for higher α than $E^H(x(\omega^c), x(\omega^c))$:

$$\begin{aligned} E^H(x(\omega^c), x(\omega^c), t(\omega^c), t(\omega^c)) = 0 \iff \\ (\overline{\gamma} - \underline{\gamma})\alpha = \underline{x}/\overline{t}((1-\overline{t})(\overline{\lambda} - \underline{\lambda}) + \overline{t}(\underline{\gamma}\overline{\lambda}(1-\mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi})-1)) \\ E^L(x(\omega^n), x(\omega^n)) = 0 \iff \\ (\overline{\gamma} - \underline{\gamma})\alpha = \overline{x}/\underline{t}((1-\underline{t})(\overline{\lambda} - \underline{\lambda}) + \underline{t}(\underline{\gamma}\overline{\lambda}(1-\mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi})-1)) \end{aligned}$$

Taking the difference between the fourth and second line, we get:

$$\left(\frac{\overline{x}(1-\underline{t})}{\underline{t}} - \frac{\underline{x}(1-\overline{t})}{\overline{t}}\right)(\overline{\lambda} - \underline{\lambda}) + (\overline{x} - \underline{x})(\underline{\gamma}\overline{\lambda}(1-\mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi})-1))$$

which is positive since every term is positive.

Finally, for E^L, we have the following:

$$\begin{aligned} E^L(x(\omega^c), x(\omega^c), t(\omega^c), t(\omega^c)) = 0 \iff \\ (\overline{\gamma} - \underline{\gamma})\alpha = \underline{x}/\overline{t}(-(1-\overline{t})(\overline{\lambda}\underline{\psi} - \underline{\lambda}\overline{\psi}) + \overline{t}(\underline{\gamma}\overline{\lambda}(1-\mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi})-1)) \\ E^L(x(\omega^n), x(\omega^n)) = 0 \iff \\ (\overline{\gamma} - \underline{\gamma})\alpha = \overline{x}/\underline{t}(-(1-\underline{t})(\overline{\lambda}\underline{\psi} - \underline{\lambda}\overline{\psi}) + \underline{t}(\underline{\gamma}\overline{\lambda}(1-\mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi})-1)) \end{aligned}$$

Taking the difference between the fourth and the second line, we receive the following:

$$\left(\frac{\overline{x}(1-\underline{t})}{\underline{t}} - \frac{\underline{x}(1-\overline{t})}{\overline{t}}\right)(\underline{\lambda}\overline{\psi} - \overline{\lambda}\underline{\psi}) + (\overline{x} - \underline{x})(\underline{\gamma}\overline{\lambda}(1-\mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi})-1))$$

The inequality of the Lemma describes the case when this difference is positive. If that holds, then $E^L(x(\omega^c), x(\omega^c), t(\omega^c), t(\omega^c))$ becomes 0 for lower α than $E^H(x(\omega^n), x(\omega^n), t(\omega^n), t(\omega^n))$.

From all of the above, it follows that when condition C.2 holds, then:

$$\begin{aligned} E^L(x(\omega^c), x(\omega^c), t(\omega^c), t(\omega^c)) < E^L(x(\omega^c), x(\omega^n), t(\omega^c), t(\omega^n)) \\ < E^H(x(\omega^c), x(\omega^n), t(\omega^c), t(\omega^n)) < E^H(x(\omega^n), x(\omega^n), t(\omega^n), t(\omega^n)), \end{aligned}$$

as required. □

Note that the left-hand side of condition (C.2) is positive, so this is satisfied if the right-hand side is small enough, i.e. $\overline{\lambda}\underline{\psi} - \underline{\lambda}\overline{\psi}$ close to zero.

Assuming that condition (C.2) is satisfied, we can now characterize the equivalent of Proposition 1 for the case of endogenous taxes.

Proposition C.1. *Assume that condition C.2 holds, and that condition C.1 is satisfied for the range of α considered. Then, there are three ranges of α with cut-off values $\underline{\alpha}' < \overline{\alpha}'$ with the same properties as in Proposition 1, except that neoliberalism now comes with a tax cut. Specifically:*

1. *if and only if $\alpha \in (\underline{\alpha}', \overline{\alpha}')$, there exists an equilibrium with narrative and policy divergence where the high (low) income citizens believe in the neoliberal narrative (post-war consensus narrative), and the high (low) income party delivers deregulation (regulation) and low taxes (high taxes) if elected;*
2. *if $\alpha \leq \underline{\alpha}'$ there exists an equilibrium with full consensus on the neoliberal narrative, with deregulation and low taxes;*
3. *if $\alpha \geq \overline{\alpha}'$, there exists an equilibrium with full consensus on the post-war consensus narrative, with regulation and high taxes.*

Proof. Define $\underline{\alpha}'$ and $\overline{\alpha}'$ analogously to $\underline{\alpha}$ and $\overline{\alpha}$ for the case of exogenous taxes in Proposition 1. If C.2 and C.1 hold, then $t_J(\omega^c) = \bar{t}$ and $x_J(\omega^c) = \underline{x}$, while $t_J(\omega^n) = \underline{t}$ and $x_J(\omega^n) = \overline{x}$. The proof is then analogous to that of Proposition 1, with the exception that taxes are a function of narratives. □

The logic is the same as for Proposition 1, with the complication that some additional assumptions need to be satisfied for the analysis. In particular, it is a strong assumption to impose C.1 for all α considered while simultaneously describing equilibria as a function of α. It is necessary to find parameter combinations such that the relevant assumptions are in fact satisfied for a relevant range of α. This is what the next Proposition does, extending Proposition 1 to show under which conditions neoliberalism can lead to tax cuts:

Proposition C.2. *For high enough $\overline{\gamma}$ and low enough $\underline{\lambda}, \underline{\psi}, \underline{\gamma}$, there are three ranges of α, parallel to Proposition 1, with a consensus on neoliberalism (the post-war consensus) when α is low (high), and a middle range of α with narrative divergence. In all cases, the neoliberal (post-war consensus) narrative is associated with a preference for deregulation (regulation) and low (high) taxes.*

Proof. We look at equilibria where narratives lead to a consensus in both tax policies and regulation. We do the analysis in three steps. First, we specify the conditions required so that for each party $J \in \{L, H\}$, $x_J(\omega^n) = \overline{x}$, $x_J(\omega^c) = \underline{x}$ $t_J(\omega^n) = \underline{t}$ and $t_J(\omega^c) = \bar{t}$. Second, we specify the condition under which the

ordering of E^J functions is preserved in this case, relative to Proposition 1. Third, we show that all relevant conditions can be simultaneously satisfied for the relevant range of α when $\overline{\gamma}$ is high enough and $\underline{\lambda}, \underline{\psi}, \underline{\gamma}$ are low enough.

Tax and deregulation implied by narratives We start by finding conditions such that belief in ω^n (ω^c) induces each party to support low taxes and deregulation (high taxes and regulation). For this purpose we consider the derivatives $\frac{\partial U^J(x,t,\omega)}{\partial x}$ and $\frac{\partial U^J(x,t,\omega)}{\partial t}$ for each $J \in \{L, H\}$.

Note that the derivative in x is strictly larger for party H, and the derivative in t is strictly larger for party L. Hence, if party H (L) prefers higher t and lower x under ω^c (lower t and higher x under ω^n), the other party must prefer it as well. It follows that the binding conditions are on party H for ω^c and on party L for ω^n. Note further that $\frac{\partial^2 U^H(x,t,\omega^c)}{\partial x \partial t} = \underline{\lambda}(\overline{\gamma}(1-\mu(1+\overline{\psi}))-1) < 0$ and $\frac{\partial^2 U^L(x,t,\omega^n)}{\partial x \partial t} = \overline{\lambda}(\underline{\gamma}(1-\mu(1+\underline{\psi}))+\underline{\psi}) > 0$. Hence, for monotonicity across the domain, we require:

$$\frac{\partial U^H(x,\underline{t},\omega^c)}{\partial x} \leq 0 \iff \underline{\lambda}\left((1-\underline{t}) - \overline{\gamma}\underline{t}(\mu(1+\overline{\psi})-1)\right) \leq 0 \tag{C.3}$$

$$\frac{\partial U^L(x,\underline{t},\omega^n)}{\partial x} \geq 0 \iff \overline{\lambda}\left(\underline{\gamma}\underline{t}(1-\mu(1+\underline{\psi})) - (1-\underline{t})\underline{\psi}\right) \geq 0 \tag{C.4}$$

$$\frac{\partial U^H(\overline{x},t,\omega^c)}{\partial t} \geq 0 \iff \overline{\gamma}\alpha - \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi})-1)\overline{x} - (\alpha^H + \underline{\lambda}\overline{x}) \geq 0 \tag{C.5}$$

$$\frac{\partial U^L(\overline{x},t,\omega^n)}{\partial t} \leq 0 \iff \underline{\gamma}\alpha + \underline{\gamma}\overline{\lambda}(1-\mu(1+\underline{\psi}))\overline{x} - (\alpha^L - \underline{\psi}\overline{\lambda}\overline{x}) \leq 0 \tag{C.6}$$

Conditions C.3 and C.4 are satisfied for high enough $\overline{\gamma}$ and low enough $\underline{\psi}$, respectively. For any given α, condition C.5-C.6 are also satisfied for high enough $\overline{\gamma}$ and low enough $\underline{\gamma}$, respectively. For our analysis, we will find an α such that they are also satisfied in the relevant range.

Ordering of E^J functions With endogenous taxes, the shape of E^J functions changes, as they become functions of both regulation policies and tax policies. Lemma C.1 gives conditions under which the same ordering as for Proposition 1 is preserved. In particular, the lemma shows that the following condition needs to be satisfied:

$$\begin{aligned}&(\overline{x}-\underline{x})\left(\underline{\gamma}\overline{\lambda}(1-\mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi})-1)\right)\\ &> \left(\overline{x}(1-\underline{t})/\underline{t} - \underline{x}(1-\overline{t})/\overline{t}\right)\left(\overline{\lambda}\underline{\psi} - \underline{\lambda}\overline{\psi}\right)\end{aligned} \tag{C.7}$$

which is satisfied for high enough $\overline{\gamma}$ or low enough $\underline{\psi}$.

Checking assumptions for the relevant range Analogously to cut-offs in Proposition 1, we define $\underline{\alpha}'$ as the α that satisfies $E^L(x(\omega^c), x(\omega^n), t(\omega^c), t(\omega^n)) = 0$, and $\overline{\alpha}'$ as the α that satisfies $E^H(x(\omega^c), x(\omega^n), t(\omega^c), t(\omega^n)) = 0$.

Lemma C.1 shows that $\underline{\alpha}' < \overline{\alpha}'$. Moreover, it shows that if condition C.7 holds, then both these cutoffs lie above $\tilde{\alpha}'$ which is defined as the α that satisfies $E^L(x(\omega^c), x(\omega^c), t(\omega^c), t(\omega^c)) = 0$. By definition:

$$\tilde{\alpha}' = \frac{\underline{x}}{\overline{t}(\overline{\gamma} - \underline{\gamma})} \left(-(1-\overline{t})(\overline{\lambda}\underline{\psi} - \underline{\lambda}\overline{\psi}) + \overline{t}(\underline{\gamma}\overline{\lambda}(1-\mu(1+\underline{\psi})) + \overline{\gamma}\underline{\lambda}(\mu(1+\overline{\psi})-1))\right)$$

Proposition C.1 shows that cutoffs $\underline{\alpha}'$ and $\overline{\alpha}'$ determine equilibria just as in Proposition 1, as long as conditions (C.3)-(C.7) hold in the range that includes these cutoffs. We will now find conditions under which conditions (C.3)-(C.7) hold at $\tilde{\alpha}'$ as well as at values above, including $\underline{\alpha}'$ and $\overline{\alpha}'$.

First, note that conditions (C.3),(C.4) and (C.7) are independent of α. They are satisfied as long as $\overline{\gamma}$ is high enough and $\underline{\psi}$ is low enough.

For conditions (C.5) and (C.6), we first note that when $\overline{\gamma} > 1$ and $\underline{\gamma} < 1$, if these conditions are satisfied for some α^L and α^H, they will be satisfied for all higher $\alpha^L + k$, $\alpha^H + k$ with $k > 0$, too.[30] Hence, it is sufficient to show that the conditions can be satisfied for $\tilde{\alpha}'$, as the relevant cut-offs $\underline{\alpha}'$ and $\overline{\alpha}'$ lie above it.

To find combinations of parameters that satisfy all five conditions, we first note that when $\underline{\psi} = 0$ and $\underline{\lambda} = 0$, the five conditions become:

$$0\left((1-\underline{t}) - \overline{\gamma}\underline{t}(\mu(1+\overline{\psi})-1)\right) \leq 0 \qquad \text{(A.1 limit)}$$

$$\overline{\lambda}\left(\underline{\gamma}\underline{t}(1-\mu)\right) \geq 0 \qquad \text{(A.2 limit)}$$

$$\overline{\gamma}\alpha - \alpha^H \geq 0 \qquad \text{(A.3 limit)}$$

$$\underline{\gamma}\alpha + \underline{\gamma}\overline{\lambda}(1-\mu)\overline{x} - (\alpha^L) \leq 0 \qquad \text{(A.4 limit)}$$

$$(\overline{x} - \underline{x})\left(\underline{\gamma}\overline{\lambda}(1-\mu)\right) \geq 0 \qquad \text{(A.5 limit)}$$

Note that Assumptions (A.1 limit), (A.2 limit) and (A.5 limit) are now (weakly) satisfied for all parameter values. Assumption (A.3 limit) is satisfied for $\overline{\gamma} \geq \frac{1}{1-\mu}$, for any α.

Finally, we need to make sure that (A.4 limit) is satisfied in the relevant range, i. e. at and above $\tilde{\alpha}'$. With $\underline{\psi} = 0$ and $\underline{\lambda} = 0$ we have that $\tilde{\alpha}'$ becomes $\frac{\underline{x}}{(\overline{\gamma}-\underline{\gamma})}\underline{\gamma}\overline{\lambda}(1-$

[30] Increasing both α^L and α^H by k in condition (C.5) increases the LHS by $k(\overline{\gamma} - 1)$, and decreases the LHS in condition (C.6) by $k(1 - \underline{\gamma})$.

μ).[31] Plugging this in for condition C.6, we then require that

$$\underline{\gamma}\left(\frac{\underline{\gamma}\underline{x}}{\overline{\gamma}-\underline{\gamma}}\overline{\lambda}(1-\mu)+\overline{\lambda}(1-\mu)\overline{x}\right)\leq\tilde{\alpha}^{L}$$

where $\tilde{\alpha}'=\mu\tilde{\alpha}^{L}+(1-\mu)\tilde{\alpha}^{H}$.

Since the left-hand side is increasing in $\underline{\gamma}$, this will be satisfied for low enough $\underline{\gamma}$. Hence, when $\underline{\gamma}$ is low enough to satisfy the above condition, $\underline{\psi}=\underline{\lambda}=0$, and $\overline{\gamma}\geq\frac{1}{1-\mu}$, all five conditions are satisfied at the cut-off $\tilde{\alpha}'$ as well as at higher $\underline{\alpha}'$ and $\overline{\alpha}'$.

□

This result shows that the logic of the model holds over a two-dimensional policy choice where a belief in neoliberalism now comes with low taxes and deregulation. Intuitively, this is because the low $\underline{\gamma}$ associated with the neoliberal narrative implies that public goods provision is inefficient, favoring low taxes. Hence, the same broad conclusions can be drawn with endogenous taxes as in Proposition 1. Specifically, divergence in equilibrium narratives and policies, which now include both taxation and regulation, occur for intermediate values of α. This is a key takeaway from the model of endogenous narratives, and plays a particular role in the dynamic analysis in section 4.4 of the main text.

Alternative trickle-down effects

At the core of our analysis lies a "predistributive" policy x that we have referred to as regulation, which increases the private income of the rich while decreasing that of the poor. Despite this divergence in effects of x on income, our model gives rise to situations where citizens of both income groups converge on narratives that support high or low x, depending on parameter values and the state of the economy as represented by α. Given that x directly reduces the incomes of low-income citizens, they will only adopt a subjective causal model associated with high x if there is an offsetting benefit. Thus, a trickle-down mechanism that compensates the poor for the direct loss in income is essential to reach a consensus on a neoliberal narrative that favors high x. Similarly, any model that allows for a consensus on narratives in support of low x will require a mechanism that

[31] Note also that $\underline{\alpha}'$ will lie strictly between $\frac{\underline{x}}{(\overline{\gamma}-\underline{\gamma})}\underline{\gamma}\overline{\lambda}(1-\mu)$ and $\frac{\overline{x}}{(\overline{\gamma}-\underline{\gamma})}\underline{\gamma}\overline{\lambda}(1-\mu)$, which is above $\tilde{\alpha}'$. At the same time, $\overline{\alpha}'$ will lie strictly between $\frac{\underline{x}}{\overline{t}(\overline{\gamma}-\underline{\gamma})}((1-\overline{t})\overline{\lambda}+\overline{t}\underline{\gamma}\overline{\lambda}(1-\mu))$ and $\frac{\overline{x}}{\underline{t}(\overline{\gamma}-\underline{\gamma})}((1-\underline{t})\overline{\lambda}+\underline{t}\underline{\gamma}\overline{\lambda}(1-\mu))$, strictly above $\underline{\alpha}'$.

compensates high-income voters for the loss of income associated with regulation.

The trickle-down mechanism we have proposed in our model runs through the provision of public goods and generates a feedback effect from the aggregate economy to the payoffs of citizens in either income group. In particular, this trickle-down effect encourages the poor to adopt neoliberal narratives if the loss to private incomes from high x is low compared to the gain in public output, which happens when α is low. Similarly, it discourages high-income citizens from adopting neoliberal narratives when the public good loss of x is too high, which happens for high α.

Our model highlights the importance of neoliberal scepticism towards state effectiveness in shaping equilibria. If the public sector is wasteful and inefficient, then collective gains from higher incomes are squandered if they are taxed away by the state. This generates a trade-off between private and public returns to deregulation which fuels support for neoliberal narratives.

There are other ways to generate a trickle-down mechanism that can also create a similar trade-off. For the state of the economy, represented by α, to affect the attractiveness of neoliberalism, there would need to be a role for the aggregate effects of deregulation to matter so that low-income (high-income) citizens might be convinced to support high (low) x.

One way to model this is to think of post-war consensus or neoliberalism as reflecting views about collectivism or individualism. Neoliberal narratives then focus on self-interested benefits when forming policy preferences while the post-war consensus stresses societal benefits. A simple way to capture this idea, without appealing to provision of public goods, is to define a new parameter, $\tilde{\gamma}$, that parameterizes collectivism when it multiplies the *average* societal utility in citizens' payoffs. As a causal parameter, it tries to convince citizens that increasing average utility leads to higher payoffs to both groups of citizens. A higher value of $\tilde{\gamma}$ now represents a more collectivist worldview, while a lower $\tilde{\gamma}$ is a stronger form of individualism. If we leave all other aspects of the model as above, private incomes are still given by equation (6) and the role of parameters λ and ψ are unaltered with deregulation x still lowering the private incomes of the poor while increasing those of the rich.

The utility of an individual from income group K with narratives $(\psi, \lambda, \tilde{\gamma})$ is then given by:

$$U^K(t, x; \psi, \lambda, \tilde{\gamma}) = y^K(x; \psi, \lambda) + \tilde{\gamma} Y(x; \psi, \lambda). \tag{C.8}$$

Comparing this with equation (7) of the original model, we see that utility is the same, except that $\tilde{\gamma}$ replaces γ, private income y^K is not multiplied by $(1-t)$ while aggregate income Y is no longer multiplied by t. It should be clear that the two interpretations of the difference between the post-war consensus and neoliberalism are now formally very similar; indeed, they are exactly equivalent when $t = 1 - t = 0.5$. Thus, all of our previous propositions would go through under this alternative formulation of the trickle-down effect.

This offers an alternative interpretation of neoliberal narratives and their adoption. Even without disagreement on government efficiency, the post-war consensus narrative ω^c and neoliberalism ω^n offer alternative views about the importance of social solidarity as represented by $\tilde{\gamma}$, where the former embraces a collectivist and the latter an individualistic approach. This confirms the view that World War II and its narrative of collective sacrifice helped to build the post-war consensus but that this fell apart during the economic turbulence of the 1970s. Specifically, this version of events says that high-income citizens were willing to subscribe to the post-war consensus as long as α was high enough, because they valued high collective utility. However, once α fell below $\overline{\alpha}$ (see Proposition 1), this broke the consensus with high-income citizens drawn to the individualism implicit in neoliberal narratives that encouraged them to focus on their individual payoffs, and support a high x. If α falls below $\underline{\alpha}$, the benefits of high x to higher-income people even convince those of lower incomes to adopt neoliberal narratives to share in the higher aggregate economic payoffs generated by high x.[32]

[32]See Shayo (2009) for a similar mechanism in the context of nationalism.