A Model of Competing Narratives*

Kfir Eliaz and Ran Spiegler†

Previous version: January 2019
This version: July 2019

Abstract

We formalize the argument that political disagreements can be traced to a "clash of narratives". Drawing on the "Bayesian Networks" literature, we represent a narrative by a causal model that maps actions into consequences, weaving a selection of other random variables into the story. Narratives generate beliefs by interpreting long-run correlations between these variables. An equilibrium is defined as a probability distribution over narrative-policy pairs that maximize a representative agent's anticipatory utility - capturing the idea that people are drawn to hopeful narratives. Our equilibrium analysis sheds light on the structure of prevailing narratives, the variables they involve, the policies they sustain and their contribution to political polarization.

*Financial support by ERC Advanced Investigator grant no. 692995 is gratefully acknowledged. We thank Yotam Alexander, Alessandra Cassela, Elhanan Helpman, Ariel Rubinstein, Heidi Thysen, Stephane Wolton, as well as seminar and conference audiences at Bonn, Haifa, Penn, Princeton, Johns Hopkins, Boston College, University, ESSET and CCET for helpful comments.

†Eliaz: School of Economics, Tel-Aviv University and David Eccles School of Business, University of Utah. E-mail: kfire@tauex.tau.ac.il. Spiegler: School of Economics, Tel-Aviv University and Economics Dept., University College London and CFM. E-mail: rani@post.tau.ac.il.

1 Introduction

The idea that political disagreements can be traced to a "*clash of narratives*" has become commonplace. According to this view, divergent opinions involve more than heterogeneous preferences or information: They can arise from conflicting *stories* about political reality. As a result, public-opinion makers try to shape the popular narratives that surround policy debates, because a policy gains in popularity if it can be sustained by an effective narrative.

There are countless expressions of this idea in popular and academic discourse. For instance, a recent profile of a former aide of President Obama begins with the words "Barack Obama was a writer before he became a politician, and he saw his Presidency as a struggle over narrative".[1] Likewise, two public policy professors write: "There can be little doubt then that people think narratives are important and that crafting, manipulating, or influencing them likely shapes public policy". They add that narratives simplify complex policy issues "by telling a story that includes assertions about what causes what, who the victims are, who is causing the harm, and what should be done".[2]

In this paper we offer a formalization of the idea that battles over public opinion involve competing narratives. Of course, the term "narrative" is vague and any formalization inevitably leaves many of its aspects outside the scope of investigation. Our model is based on the idea that in the context of policy debates, narratives can be regarded as *causal models* that map actions to consequences. Following the literature on probabilistic graphical models in Statistics, Artificial Intelligence and Psychology (Cowell et al. (1999), Sloman (2005), Pearl (2009)), we represent such causal models by directed acyclic graphs (DAGs).

In our model, what defines a narrative is the variables it incorporates and the way these are arranged in the causal mapping from actions to consequences. For instance, consider a debate over US trade policy and its

[1] See https://www.newyorker.com/magazine/2018/06/18/witnessing-the-obama-presidency-from-start-to-finish.

[2] See http://blogs.lse.ac.uk/impactofsocialsciences/2018/07/18/mastering-the-art-of-the-narrative-using-stories-to-shape-public-policy/.

possible implications for local employment. Suppose the public has homogenous preferences over actions and consequences; disagreements only arise from different beliefs. The DAG

$$\text{trade policy} \quad \rightarrow \quad \text{imports from China} \quad \rightarrow \quad \text{employment} \tag{1}$$

represents a narrative that weaves a third variable (imports from China) into a causal story about the employment consequences of trade policy.

The nodes in the DAG represent variables (not the values they can take), and the links represent perceived direct causal effects (but not the sign or magnitude of these effects). The variables are coarse-grained, such that the narrative does not describe an individual historical episode. Instead, it addresses numerous historical episodes, alerting the public's attention to long-run correlations between adjacent variables along the causal chain and offering a particular causal interpretation of these correlations. In general, our model assumes that when the public adopts a narrative, it constructs a belief by fitting the causal model to objective data. As in Spiegler (2016), this means factorizing the long-run distribution (over the variables that appear in the narrative) according to the Bayesian-Network factorization formula. The public then relies on this belief to evaluate policies.

We refer to the causal model (1) as a "*lever narrative*" because it regards imports from China as a "lever" (or a mediator, to use statisticians' jargon) - i.e., an endogenous variable that is influenced by policy and in turn influences the target variable. To the extent that imports from China are negatively correlated with both protectionism and employment in local manufacturing, this narrative intuitively supports a protectionist policy. But while the support is intuitive, it is illusory if the narrative is false - e.g. if the actual correlation between employment and imports from China is due to confounding by exogenous technological change. A false narrative will typically induce a distorted belief regarding the mapping from actions to consequences.

The following is another example of a lever narrative in the context of a foreign policy question, whether to impose economic sanctions on a rival country with a hostile regime. The public considers destabilizing the regime a desirable outcome. A lever narrative that intuitively gives support to a hawkish policy is

$$\text{sanction policy} \quad \rightarrow \quad \text{economic situation in rival country} \quad \rightarrow \quad \text{regime stability}$$

The following is a lever narrative that involves a different "lever":

$$\text{sanction policy} \quad \rightarrow \quad \text{nationalism in rival country} \quad \rightarrow \quad \text{regime stability}$$

This narrative intuitively supports a *dovish* policy, to the extent that nationalistic sentiments in the rival country are positively correlated with the stability of its regime and ameliorated by a soft stance on sanctions. We can see that two narratives may have the same "lever" structure but differ in the selection of variables that function as "levers", and consequently in the policies they support.

Likewise, the same variable can be assigned different roles in the causal scheme. For instance, the following is a foreign-policy narrative that treats nationalism as an *exogenous* variable:

$$\text{sanction policy} \quad \rightarrow \quad \text{regime stability} \quad \leftarrow \quad \text{nationalism in rival country}$$

We refer to a narrative with this structure as a "threat/opportunity narrative", because it regards the third variable that it weaves into the story as an external factor that the policy takes into consideration rather than influencing it. In Section 3.1 we will show how this narrative can lend support to a *hawkish* policy.

Thus, narratives can differ in the variables they involve or in the role that these variables play in the causal mapping from actions to consequences. Different narratives can generate different political beliefs because they alert the audience's attention to correlations between different sets of variables and shape the audience's causal interpretation of these correlations.

But how does the public respond to competing narratives that support conflicting policies? In the context of policy debates, we find it natural to assume that people are drawn to *hopeful* narratives. By "hopeful", we do not mean that appealing narratives portray a rosy picture of the status quo, but rather that they are expected to promise a "better future" if a certain policy is implemented. Precisely because individuals have little influence over public policy, they incur negligible decision costs when indulging in hopeful fantasies about the effects of counterfactual policies. Therefore, anticipatory feelings can be a powerful driving force behind political positions.[3]

Accordingly, we assume that the public selects a narrative-policy that maximizes *anticipatory utility*, subject to one empirical-consistency constraint: To be eligible, a narrative is not allowed to distort the steady-state distribution over consequences. In other words, narratives can spin hopeful fantasies about the consequences of the policies they espouse, but not about the status quo.

To sum up, our model is based on two related premises. First, political beliefs are shaped by narratives, which are simplified causal models that interpret long-run correlations. Second, in the presence of competing narratives, people are drawn to the ones that promise a "happy ending". On these premises, we define an *equilibrium* as a long-run distribution over narrative-policy pairs, such that every element in the support maximizes a representative agent's anticipatory utility, subject to the above empirical-consistency constraint.

We refer to this concept as "equilibrium" rather than mere optimization. The reason is that the distribution over policies can affect the way a narrative-induced belief evaluates individual policies. As a result, a change in the policy distribution can lead to a change in the relative appeal of competing narratives. This feedback effect is a hallmark of behavior that is generated by misspecified causal models (see Spiegler (2016)), and it is what creates the need for an equilibrium approach to the notion of prevalent narratives.

We employ our equilibrium concept to explore several questions: What is the structure of narratives that support a given policy, and what kind

[3]In their book on the use of narratives to win public support, De Graaf et al. (2015) argue that one of the major features of an effective narrative is the prospect of success: "... the overarching story told by incumbent policy-makers must, to some extent, be a narrative of progress."

of variables do they involve? Can we account for divergent political beliefs or swings between dominant positions? Can we shed light on the source of popularity of certain real-life political narratives? Our results demonstrate the formalism's potential to shed light on such questions.

Related literature

The idea that people think about empirical regularities in terms of "causal stories" that can be represented by DAGs has been embraced by psychologists of causal reasoning (e.g. Sloman (2005), Sloman and Lagnado (2015)). Spiegler (2016) adopted this idea as a basis for a model of decisions under causal misperceptions, in which the decision maker forms a subjective belief by fitting a subjective causal model to objective long-run data. This continues to be a building block of the model in this paper, which goes beyond it in two major directions. First, the collection of variables that can appear in a causal model of a given size is not fixed but selected endogenously. Second, we assume "hedonic" selection between competing causal models.

We are aware of at least three papers in economics that draw attention to the role of narratives in economic contexts. Given that the term "narrative" has such a loose meaning, it should come as no surprise that it has received very different formalizations. Shiller (2017) regards certain terms that appear in popular discourse as indications of specific narratives and proposes to use epidemiological models to study their spread. Benabou et al. (2016) focus on moral decision making and formalize narratives as messages or signals that can affect decision makers' beliefs regarding the externality of their actions. Levy and Razin (2018) use the term to describe information structures in game-theoretic settings that people postulate to explain observed behavior. Schwartzstein and Sunderam (2019) propose an alternative approach to "persuasion by models", where models are formalized as likelihood functions and the criterion for selecting models is their success in accounting for historical observations.

The idea that people adopt distorted beliefs to enhance their anticipatory utility has precedents in the economics literature: Akerlof and Dickens (1982), Benabou and Tirole (2002,2016), Brunnermeier and Parker (2005) and Spiegler (2008). Relative to this literature, the key innovation here is that the object of agents' choice is not beliefs but (causal) *models*: Wrong beliefs emerge as a consequence of fitting a misspecified model to historical data. This feature constrains agents' ability to delude themselves and leads to novel equilibrium effects. Recently, Montiel Olea et al. (2018) studied "competing models" in a different context of experts who compete for the right to make predictions. Each expert believes in a linear regression model that differs in the set of variables it admits. Winning models thus maximize the indirect expected utility they induce when estimated against a random sample.

Finally, our paper joins a handful of works in so-called "behavioral political economics" that study voters' belief formation according to misspecified subjective models or wrong causal attribution rules - e.g., Spiegler (2013), Esponda and Pouzo (2017).

2 The Model

Let $X = X_1 \times \cdots \times X_m$, where $m > 2$ and $X_i = \{0, 1\}$ for each $i = 1, ..., m$. For every $N \subseteq \{1, ..., m\}$, denote $X_N = \times_{i \in N} X_i$. For any $x \in X$, the components x_1 and x_m - also denoted a and y - are referred to as the *action* and the *consequence*. Let $p \in \Delta(X)$ be an objective probability distribution with full support. Denote $p(a = 1) = \alpha$ and $p(y = 1) = \mu$. We interpret α as a historical, long-run action frequency and endogenize it later in this section.

A *directed acyclic graph* (DAG) is a pair $G = (N, R)$, where $N \subseteq \{1, ..., m\}$ is a set of nodes and $R \subseteq N \times N$ is a set of directed links. Acyclicity means that the graph contains no directed path from a node to itself. We use iRj or $i \rightarrow j$ to denote a directed link from the node i into the node j. Abusing notation, let $R(i) = \{j \in N \mid jRi\}$ be the set of "parents" of node i. Following Pearl (2009), we interpret a DAG as a *causal model*, where the link $i \rightarrow j$ means that x_i is perceived as an immediate cause of x_j. Directedness and acyclicity of G are consistent with basic intuitions regarding causality. The causal model is agnostic about the sign or magnitude of causal effects.

Let $\mathcal{G}$ be a collection of DAGs. We refer to an element in $\mathcal{G}$ as a *narrative*. Every $G \in \mathcal{G}$ satisfies the following restrictions. First, $\{1, m\} \subseteq N$, capturing the idea that the narratives concern the consequences of policy. Second, $|N| \leq n$, where $n \in \{2, ..., m\}$ is an exogenously given constant that represents an upper bound on narrative complexity. Third, 1 is an *ancestral* node. This restriction means that actions have no prior causes. We relax this restriction in Section 5. In applications, we will impose additional restrictions on $\mathcal{G}$ because certain causal models are implausible on a-priori grounds (e.g. assuming that tariff policy has no causal effect on imports), or because the data they require for generating beliefs (according to the definition we provide in the next paragraph) is unavailable.

From narratives to beliefs
Given an objective distribution p, a narrative $G = (N, R)$ induces a subjective belief over $\Delta(X_N)$, defined as follows:

$$p_G(x_N) = \prod_{i \in N} p(x_i \mid x_{R(i)}) \tag{2}$$

The full-support assumption ensures that all the terms in this factorization formula are well-defined.

The conditional distribution of x_m given x_1 induced by p_G is computed in the usual way. It has a simple expression because 1 is an ancestral node:

$$p_G(x_m \mid x_1) = \sum_{x_{N-\{1,m\}}} \left(\prod_{i \in N-\{1\}} p(x_i \mid x_{R(i)}) \right) \tag{3}$$

The fact that 1 is ancestral also ensures that this conditional distribution has a natural interpretation as the perceived causal effect of x_1 on x_m, according to the belief p_G.

For illustration, when $n = m = 4$ and the narrative is $G : 1 \rightarrow 3 \rightarrow 4 \leftarrow 2$,

$$p_G(x_1, x_2, x_3, x_4) = p(x_1)p(x_2)p(x_3 \mid x_1)p(x_4 \mid x_2, x_3)$$

and

$$p_G(x_4 \mid x_1) = \sum_{x_2, x_3} p(x_2)p(x_3 \mid x_1)p(x_4 \mid x_2, x_3)$$

The induced marginal distribution over consequences is

$$p_G(x_m) = \sum_{x_1,...,x_{m-1}} p_G(x_1, ..., x_m) \tag{4}$$

Formula (2) is the standard Bayesian-network factorization formula (see Spiegler (2016) and the references therein). Its interpretation in the current context is as follows. A narrative selects up to $n-2$ variables (other than the action and the consequence) and incorporates them into a causal story. This is akin to a novelist who conjures up a collection of events, and then organizes their unfolding according to a plot. The narrative generates a subjective belief regarding the mapping from actions to consequences, by drawing the audience's attention to particular correlations (those deemed relevant by the causal model) and combining them in accordance with the causal model. The correlations themselves are accurate - i.e., each of the terms in the factorization (2) is extracted from the objective distribution p. It is the way they are combined that may lead to distorted beliefs.

Policies and anticipatory utility

Let $D = [\varepsilon, 1-\varepsilon]$, where $\varepsilon > 0$ is arbitrarily small. A *policy* $d \in D$ is a proposed frequency of playing the action $a = 1$.[4] A representative agent has a utility function $u(y, d) = y - C(d - d^*)$, where $d^* \in D$ is the agent's ideal policy, and C is a symmetric, convex cost function that satisfies $C(0) = C'(0) = 0$. Thus, $y = 1$ is the agent's desirable outcome, and the function C represents the intrinsic disutility he experiences when deviating from his ideal policy.

Given p, a narrative-policy pair (G, d) induces *gross anticipatory utility*

$$V(G, d; \alpha) = d \cdot p_G(y = 1 \mid a = 1) + (1-d) \cdot p_G(y = 1 \mid a = 0) \tag{5}$$

This is simply the subjective probability of the good outcome $y = 1$ under the policy d, according to p_G. The agent's net anticipatory utility from the

[4] We define a policy as a mixture over actions rather than identifying it with a, in order to prevent certain interesting effects from being obscured or trivialized.

narrative-policy pair (G, d) given p is

$$U(G, d; \alpha) = V(G, d; \alpha) - C(d - d^*) \tag{6}$$

The notation $V(R, d; \alpha)$ highlights a crucial feature: A change in α (namely the marginal of p over a) can alter $p_G(y \mid a)$, and therefore $V(G, d; \alpha)$. This would be impossible under rational expectations, as $p(y \mid a)$ is invariant to α *by definition*. We will see this effect in action in Section 3.

Recall that we require a to be an ancestral node in G. As a result, $p_G(a = 1) = \alpha$ (see Spiegler (2017)). This means that

$$\begin{aligned} V(G, \alpha; \alpha) &= p_G(a = 1) \cdot p_G(y = 1 \mid a = 1) + p_G(a = 0) \cdot p_G(y = 1 \mid a = 0) \\ &= p_G(y = 1) \end{aligned} \tag{7}$$

In other words, the gross anticipatory utility from a "status quo" policy that mimics the objective long-run action frequencies is necessarily equal to the marginal distribution over consequences implied by p_G.

Equilibrium
The model's exogenous components are the conditional distribution $(p(x_2, ..., x_m \mid x_1))$, the set of feasible narratives $\mathcal{G}$ and the cost function C. We are now ready to define our notion of equilibrium, which endogenizes α.

Definition 1 *An action frequency $\alpha \in [0, 1]$ and a probability distribution σ over narrative-policy pairs (G, d) constitute an equilibrium if:*
(i) Any $(G, d) \in Supp(\sigma)$ maximizes $U(G, d; \alpha)$ subject to $V(G, \alpha; \alpha) = \mu$.
(ii) $\alpha = \sum_{(G,d)} \sigma(G, d) \cdot d$

This solution concept captures a steady state in the battle over public opinion. The first condition requires that prevailing narrative-policy pairs are those that maximize the representative agent's net anticipatory utility, subject to a constraint we dub "*No Status-Quo Distortion*" (NSQD). Thus, the public's criterion for selecting between competing narrative-policy pairs is net anticipatory utility. This captures the idea that voters do not adjudicate between narratives using "scientific" methods; rather, they are attracted to narratives with a "hopeful" message (stories with a happy ending, so to speak). When faced with contradictory causal models, the public behaves as if it believes, Paraphrasing George Box's famous quote, that "all models are wrong, but some are hopeful".[5]

The NSQD constraint

Viewed formally, NSQD is the familiar Bayes plausibility condition: The expected posterior distribution over y should coincide with its marginal distribution. In the present context, it can be interpreted as follows. Suppose that the narrative G is adopted. Then it is natural for the audience to contemplate its implications when coupled with the status-quo policy α. If $V(G, \alpha; \alpha) \neq \mu$, it is as if the narrative makes the absurd statement: "Let us keep doing what we have done so far, and the outcome would be different". NSQD rules out such narratives: It allows narratives to make false promises about *counterfactual* policies, but it does not allow them to distort the status quo.

By (7), NSQD is equivalent to requiring $p_G(y = 1) = \mu$. This enables a more direct interpretation of NSQD as an empirical-consistency criterion that constrains narratives' ability to delude the public: Prevailing narratives are not allowed to induce beliefs that distort the steady-state distribution over consequences. The justification is that while testing correlations between variables is a difficult task for voters, monitoring the long-run behavior of the target variable y is relatively easy. Therefore, it will be relatively easy to discredit a narrative that induces a belief that is inconsistent with the long-run observations of y.

In certain cases, the DAG's structure alone ensures NSQD. A DAG (N, R) is *perfect* if whenever iRk and jRk for some $i, j, k \in N$, it is the case that iRj or jRi. Thus, in a perfect DAG, if two variables are perceived to be direct causes of a third variable, then there must be a perceived direct causal link

[5]Dahlstrom (2013) writes that "narratives can also perpetuate misinformation...accepted narratives are trusted so much that individuals rarely allow evidence to contradict the narrative; evidence is altered to fit their narratives." McComas and Shanahan (1999) and Szostek (2018) also argue that people's attachment to a particular narratives is not necessarily based on scientific scrutiny.

between them. E.g., $1 \to 2 \to 3$ is perfect, whereas $1 \to 3 \leftarrow 2$ is imperfect. Perfection is a familiar property in the Bayesian Networks literature. In our context, the crucial property of perfection is its relation to NSQD. If G is perfect, then $V(G, \alpha; \alpha) = \mu$ for every objective distribution p with any given α, μ. Conversely, if G is imperfect, we can find objective distributions for which NSQD fails. This result is stated and proved in Spiegler (2018) in a different context. Thus, from now on, we only need to check NSQD for imperfect DAGs.

Condition (ii) in Definition 1 endogenizes the historical action frequency α by requiring it to be consistent with the marginal of σ over policies (the lower and upper limits of D ensure that α is interior). We offer two interpretations of Condition (ii). A static, "cross-section" interpretation is that σ describes the relative popularity of narrative-policy pairs, such that α is a popularity-weighted average policy. An alternative interpretation is "ergodic". At any point in time, a particular policy rises to dominance because its accompanying narrative appeals to the public. Over time, as the long-run action frequency gravitates toward the dominant policy, the anticipatory payoff induced by various narrative-policy pairs can change. As a result, a different narrative-policy pair can become dominant. The distribution α is the average action frequency that results from periodic swings between dominant narrative-policy pairs. As in the case of conventional solution concepts like Nash or competitive equilibrium, this interpretation raises the question of whether it can be backed by an explicit dynamic mechanism. We will return to this question in the sequel.

The following is a simple rational-expectations benchmark. Suppose that $\mathcal{G}$ consists of a single narrative $G : a \to y$. Then, $p_G(y \mid a) \equiv p(y \mid a)$. In equilibrium (α, σ), σ assigns probability one to policies d that maximize $d \cdot p(y = 1 \mid a = 1) + (1 - d) \cdot p(y = 1 \mid a = 0) - C(d - d^*)$. When C is strictly convex, the equilibrium is unique.

From now on, we depart from this benchmark and assume that the model's primitives are minimally rich in the following sense.

Definition 2 *The pair* $(p, \mathcal{G})$ *is* non-null *if there exist* $G, G' \in \mathcal{G}$ *such that*

$p_G(y \mid a)$ *is non-constant in* a *and* $p_{G'}(y \mid a) = \mu$ *for all* a.

Thus, the set of feasible narratives is rich enough to enable a belief that actions can affect consequences, as well as a belief that the distribution of consequences is independent of actions. For instance, if $\mathcal{G}$ contains a DAG in which a and y are both ancestral nodes, as well as a DAG $a \to x_k \to y$ such that x_k is correlated with both a and y, then $\mathcal{G}$ is non-null. All the results in the paper take it for granted that $(p, \mathcal{G})$ is non-null.

Proposition 1 *An equilibrium exists.*

The proof of this result involves constructing an auxiliary game, such that existence of Nash equilibrium in this game is equivalent to existence of our notion of equilibrium.

3 Examples

In this section we illustrate our formalism with two simple examples.

3.1 Foreign-Policy Narratives

Let $m = n = 3$. The three variables are as follows. The action a represents the attitude toward a rival country having a hostile regime, where $a = 1$ (0) denotes a hawkish (dovish) attitude. The consequence y represents the hostile regime's stability, where $y = 1$ (0) indicates regime change (regime stability). The third variable, denoted s, represents nationalistic sentiments in the other country, where $s = 1$ (0) indicates strong (weak) nationalism.

The exogenous aspects of the objective distribution p are as follows. First, $p(y = 1) = \frac{1}{2}$, independently of a. The interpretation is that foreign policy has no effect on regime stability. Second, $p(s = 1 \mid a, y) = (a+1-y)/2$. Thus, nationalism is positively correlated with both hawkish policy and regime stability. However, these two correlations have different causal meaning. The correlation between a and s is causal: hawkish (dovish) policy tends to strengthen (weaken) nationalism in the other country. In contrast, the correlation between s and y is *not* causal; rather, it is due to confounding by unmodeled exogenous factors.

The set $\mathcal{G}$ consists of all DAGs that include a (as an ancestral node) and y. This set can be classified into three groups: the lever narrative $G^L : a \rightarrow s \rightarrow y$; the threat/opportunity narrative $G^O : a \rightarrow y \leftarrow s$; and all remaining narratives, which can easily be shown to induce the rational-expectations belief $p_G(y = 1 \mid a) = \frac{1}{2}$ for all a. Finally, the parameter ε that defines D is vanishingly small. The cost function is $C \equiv 0$, hence the value of d^* is immaterial.

Claim 1 *There exists a unique equilibrium* (α^*, σ^*)*, where* $\alpha^* \approx 0.57$ *and* $Supp(\sigma^*) = \{(G^L, \varepsilon), (G^O, 1-\varepsilon)\}$.

Proof. We give a detailed proof in order to demonstrate the type of calculations that analyzing our notion of equilibrium involves. Proofs of our other results are relegated to the appendices.

We first derive $p_G(y \mid a)$ for every $G \in \mathcal{G}$. Any G that induces $p_G(y = 1 \mid a) = \frac{1}{2}$ for all a generates $U(G, d; \alpha) = \frac{1}{2}$ for any d, α. Now consider the narrative G^O:

$$p_{G^O}(y = 1 \mid a) = p(s = 1)p(y = 1 \mid a, s = 1) + p(s = 0)p(y = 1 \mid a, s = 0)$$

Plugging our specification of $p(a, y, s) = p(a)p(y)p(s \mid a, y)$, we obtain

$$p(s = 1) = \alpha \cdot \frac{1}{2} \cdot 1 + (1 - \alpha) \cdot \frac{1}{2} \cdot 0 + \left[\alpha \cdot \frac{1}{2} + (1 - \alpha) \cdot \frac{1}{2}\right] \cdot \frac{1}{2} = \frac{1}{4} + \frac{\alpha}{2}$$

and

$$\begin{aligned}
p(y &= 1 \mid a=1, s=1) = \frac{\alpha \cdot \frac{1}{2} \cdot \frac{1}{2}}{\alpha \cdot \frac{1}{2} \cdot \frac{1}{2} + \alpha \cdot \frac{1}{2} \cdot 1} = \frac{1}{3} \\
p(y &= 1 \mid a=1, s=0) = \frac{\alpha \cdot \frac{1}{2} \cdot \frac{1}{2}}{\alpha \cdot \frac{1}{2} \cdot \frac{1}{2} + \alpha \cdot \frac{1}{2} \cdot 0} = 1 \\
p(y &= 1 \mid a=0, s=1) = 0 \\
p(y &= 1 \mid a=0, s=0) = \frac{(1-\alpha) \cdot \frac{1}{2} \cdot 1}{(1-\alpha) \cdot \frac{1}{2} \cdot 1 + (1-\alpha) \cdot \frac{1}{2} \cdot \frac{1}{2}} = \frac{2}{3}
\end{aligned}$$

Therefore,

$$\begin{aligned}
p_{G^O}(y &= 1 \mid a=1) = \frac{5}{6} - \frac{\alpha}{3} \\
p_{G^O}(y &= 1 \mid a=0) = \frac{1}{2} - \frac{\alpha}{3}
\end{aligned}$$

such that

$$V(G^O, d; \alpha) = d(\frac{5}{6} - \frac{\alpha}{3}) + (1-d)(\frac{1}{2} - \frac{\alpha}{3})$$

Plugging $d = \alpha$, we can confirm that $V(G^O, \alpha; \alpha) = \frac{1}{2}$ - i.e., G^O satisfies NSQD for any α. Moreover, for any α, $V(G^O, d; \alpha)$ is strictly increasing in d. Therefore, if (G^O, d) is in the support of the equilibrium, then $d = 1 - \varepsilon$. Since $V(G^O, 1-\varepsilon; \alpha) \approx \frac{5}{6} - \frac{\alpha}{3} > \frac{1}{2}$ for any $\alpha < 1$, it follows that no narrative that induces rational expectations can prevail in equilibrium.

Next, consider the narrative G^L:

$$p_{G^L}(y=1 \mid a) = p(s=1 \mid a)p(y=1 \mid s=1) + p(s=0 \mid a)p(y=1 \mid s=0)$$

Plugging our specification of p, we obtain

$$\begin{aligned}
p(s &= 1 \mid a=1) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot 1 = \frac{3}{4} \\
p(s &= 1 \mid a=0) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}
\end{aligned}$$

and

$$
\begin{aligned}
p(y &= 1 \mid s = 1) = \frac{\frac{\alpha}{2}\cdot\frac{1}{2} + \frac{1-\alpha}{2}\cdot 0}{\frac{1}{4}+\frac{\alpha}{2}} = \frac{\alpha}{1+2\alpha} \\
p(y &= 1 \mid s = 0) = \frac{\frac{\alpha}{2}\cdot\frac{1}{2} + \frac{1-\alpha}{2}\cdot 1}{\frac{3}{4}-\frac{\alpha}{2}} = \frac{2-\alpha}{3-2\alpha}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
p_{G^L}(y &= 1 \mid a = 1) = \frac{3}{4}(\frac{\alpha}{1+2\alpha}) + \frac{1}{4}(\frac{2-\alpha}{3-2\alpha}) = \frac{1+6\alpha-4\alpha^2}{2(1+2\alpha)(3-2\alpha)} \\
p_{G^L}(y &= 1 \mid a = 0) = \frac{1}{4}(\frac{\alpha}{1+2\alpha}) + \frac{3}{4}(\frac{2-\alpha}{3-2\alpha}) = \frac{3+6\alpha-4\alpha^2}{2(1+2\alpha)(3-2\alpha)}
\end{aligned}
$$

such that

$$
V(G^L, d; \alpha) = d \cdot \frac{1+6\alpha-4\alpha^2}{2(1+2\alpha)(3-2\alpha)} + (1-d)\cdot\frac{3+6\alpha-4\alpha^2}{2(1+2\alpha)(3-2\alpha)} \tag{8}
$$

Because G^L is perfect, it necessarily satisfies NSQD (as can be verified by setting $d = \alpha$ in (8)). Note that $V(G^L, d; \alpha)$ is strictly decreasing in d. Therefore, if (G^L, d) is in the support of the equilibrium, then $d = \varepsilon$.

We have seen that any narrative $G \neq G^O, G^L$ cannot prevail in equilibrium. We now show that *both* G^O and G^L belong in the equilibrium support. Assume the contrary, and suppose G^O is the only narrative in the support. Then, as shown above, it will be paired with the policy $d = 1-\varepsilon$, such that $\alpha = 1-\varepsilon$. Since G^O satisfies NSQD, $V(G^O, 1-\varepsilon; 1-\varepsilon) = \frac{1}{2}$. But since $V(G^L, \varepsilon; 1-\varepsilon) \approx \frac{5}{6}$, $(G^O, 1-\varepsilon)$ does not maximize the agent's anticipatory payoff, a contradiction. Similarly, if G^L is the only prevailing narrative, it is paired with $d = \varepsilon$, such that $\alpha = \varepsilon$ and therefore $V(G^O, 1-\varepsilon; \varepsilon) \approx \frac{5}{6}$, reaching a similar contradiction.

Thus, $Supp(\sigma)$ consists of exactly two narrative-policy pairs: (G^L, ε) and $(G^O, 1-\varepsilon)$. This means that $V(G^L, \varepsilon; \alpha) = V(G^O, 1-\varepsilon; \alpha)$, which for $\varepsilon \to 0$ can be written as

$$
\frac{3+6\alpha-4\alpha^2}{2(1+2\alpha)(3-2\alpha)} \approx \frac{5}{6} - \frac{\alpha}{3}
$$

This equation has a unique solution in $[0, 1]$, $\alpha \approx 0.57$. ■

This example has a number of noteworthy features.

Coupling of narratives and policies

In this example, s is the only variable (other than a and y) that narrators can weave into their stories. However, its location in the narrative turns out to determine the endorsed policy. The narrative that sustains a hawkish policy treats s as an exogenous factor, whereas the narrative that sustains a dovish policy treats s as a lever.

The reason that the G^L promotes dovish policies is that a and s are positively correlated whereas s and y are negatively correlated. The lever narrative combines these correlations in a causal chain $a \to x \to y$. As a result, G^L (falsely) predicts a negative indirect causal effect of a on y.

The intuition for why G^O is coupled with a hawkish policy is subtler. The specification of $p(s \mid a, y)$ means that s is a (stochastic) function of the difference $a - y$. This means that for a given s, an increase in a implies an increase in the conditional probability that $y = 1$. In reality, this effect is purely diagnostic, yet G^O treats it as causal. Moreover, G^O regards the distribution of s as independent of a. It follows that G^O (falsely) predicts a positive causal effect of a on y.

Multiple prevailing narratives

The equilibrium distribution assigns weight to *two* policies. The "cross-sectional" interpretation of this effect is political polarization: At any moment in time, there are two narrative-policy pairs that dominate public opinion. The alternative, "ergodic" interpretation can be backed by an explicit dynamic-stability argument, thanks to a "*diminishing returns*" property: $V(G^L, \varepsilon; \alpha)$ is increasing in α, whereas $V(G^O, 1 - \varepsilon; \alpha)$ is decreasing in α. That is, each narrative's ability to delude the public diminishes as the policy it endorses gets implemented more frequently. Thus, if we were to perturb α above its equilibrium level (i.e., increase the frequency of $a = 1$), then (G^L, ε) would becomes more appealing than $(G^O, 1-\varepsilon)$, and therefore the prevailing policy will be dovish for some time. This pushes α back toward its original level. A similar argument applies to downward perturbation of α.

Hawkish bias
The example treats the two actions symmetrically: $p(s \mid a = 1, y) \equiv p(s \mid a = 0, y)$ and the agent has no intrinsic preference over policies. Nevertheless, the equilibrium action frequency is biased to the right. The reason is that G^O induces a false correlation $p_{G^O}(y = 1 \mid a = 1) - p_{G^O}(y = 1 \mid a = 0) = \frac{1}{3}$, which is larger in absolute terms than the correlation $-1/((1 + 2\alpha)(3 - 2\alpha))$ induced by G^L. At $\alpha = \frac{1}{2}$, this gives G^O an advantage over G^L in terms of their induced anticipatory utility. The "diminishing returns" property described above means that to equalize the narratives' anticipatory utility, α has to be greater than $\frac{1}{2}$.

Comment: Mutual narrative refutation
Our representative agent does not reason "scientifically" about conflicting narratives. He does not actively seek correlational data to test narratives. Instead, he allows "narrators" to determine the data he pays attention to. Thus, G^L alerts him to the conditional distributions $(p(s \mid a))$ and $(p(y \mid s))$, whereas G^O alerts him to $(p(s))$ and $(p(y \mid a, s))$. The data that one narrative invokes also manages to refute the *competing* narrative. The distribution $(p(s \mid a))$ referred to by G^L shows that s and a are correlated, *contra* G^O. Likewise, the distribution $(p(y \mid a, s))$ referred to by G^O demonstrates that y and a are correlated conditional on s, *contra* G^L. How would our agent react if this mutual refutation were pointed out to him? A rational reaction would be to distrust all narratives and develop a more "scientific" belief-formation method. Yet an arguably more realistic reaction would be to conclude shruggingly that "all models are wrong" and adopt the more hopeful one - especially in the political context, where the agent has virtually no "skin in the game".[6]

[6] Consider a modified example that replaces s with *two distinct variables* with the same conditional distribution. The formal analysis would be the same. However, the two conflicting narratives can invoke different variables, such that the above mutual refutation would be infeasible.

3.2 "Easy Fix" Narratives

Demagoguery is a common feature of public-opinion formation. It often involves spinning an oversimplified description of a social problem, which attributes a complex phenomenon to a single (often spurious) cause and suggests that the problem has an easy fix. So-called "populist narratives" seem to have this characteristic. In this sub-section we present a simple example that aims to capture the tension between rational and easy-fix narratives.

Let $m = n = 4$. In addition to the action and consequence variables a and y, the two other variables are denoted s and θ. The objective distribution is consistent with the following DAG $G^* : a \rightarrow s \leftarrow \theta \rightarrow y$. Accordingly, we refer to θ as the "deep cause" of y and to s as its "symptom". Furthermore, we interpret a as a policy instrument that obviously impacts s. The fact that θ is an ancestral node means that the deep cause cannot be influenced by this instrument.

We offer two economic stories for this example. In both of them, y represents the welfare of the local working class. In one story, θ represents technological change (e.g. automation), s represents foreign trade and a represents tariff policy. In the other story, θ represents spontaneous trends in developing economies, s represents immigration and a represents immigration policy.[7]

In such contexts, the lever DAG $G^L : a \rightarrow s \rightarrow y$ can be regarded as an "easy fix" narrative: It simplifies the representation of reality by neglecting the deep cause of y, and it makes a false promise by misrepresenting s as a lever for changing y. The easy-fix narrative is false because what it addresses is a mere symptom rather than the cause of the target variable y. We assume that the set of feasible DAGs is $\mathcal{G} = \{G^*, G^L\}$.[8]

Impose the following additional structure on p. First, $p(\theta = 1) = \delta$

[7] Frum (2019) proposed that rising income and human capital in developing countries allows more individuals to migrate into first-world countries.

[8] The equilibrium policy distribution we derive is robust to expanding $\mathcal{G}$ to be the set of all DAGs in which a is an ancestral node, y is a terminal node and there is a direct link $a \rightarrow s$ (in accordance with the interpretation that a is a policy instrument that manifestly impacts s). We do not present this version of the example because its proof is considerably longer.

independently of a, where $\delta \in (0,1)$. Second, $p(y = \theta \mid \theta) = \beta > \frac{1}{2}$ for all θ, independently of a, s. Third, $p(s = 1 \mid a, \theta) \approx a + (1-a)\theta$.[9] Thus, the marginal over y is given by $\mu = \beta\delta + (1-\beta)(1-\delta)$. As usual, $p(a = 1) = \alpha$. The parameter ε is vanishingly small, and the cost function is $C(\Delta) = k\Delta^2$, where $\Delta = d - d^*$. Denote $k^* = \delta(1-\delta)(\beta - \frac{1}{2})$.

Claim 2 *Assume $k < k^*$. Then, there is a unique equilibrium (α, σ), where*

$$\alpha \approx \frac{2k^* - \delta k(1-d^*)^2}{2k^* + (1-\delta)k(1-d^*)^2}$$

and $Supp(\sigma) = \{(G^*, d^*), (G^L, 1-\varepsilon)\}$.[10]

This result exhibits a subtle interplay between rational and easy-fix narratives. The rational narrative recognizes that the policy instrument is powerless to influence the deep cause of y. Therefore, it cannot offer the illusion of a "happy ending"; the only consolation it offers is a justification for taking the intrinsically desirable policy d^*. In contrast, the easy-fix narrative misreads the correlation between s and y as a causal effect and therefore conveys an illusion that y can be improved. Yet the easy-fix narrative feeds off the rational narrative; without the latter, α would coincide with the easy-fix narrative's endorsed policy, thus robbing it of the ability to convey false hope. The narrative's appeal derives precisely from the departure of its paired policy from the status quo. For this to happen, the rational narrative *must* belong to the support. In other words, *demagoguery can only thrive if public opinion gives some room to a rational narrative.*

The equilibrium average policy α increases with β. The reason is that the appeal of G^L rests on a false interpretation of the correlation between s and y. The larger β, the stronger this correlation, and therefore the larger the

[9] To satisfy the full-support assumption, define $p(s = 1 \mid a, \theta) = (1-\rho)[a + (1-a)\theta]$, where ρ is arbitrarily close to zero.

[10] The approximation of α takes ε and ρ (see the previous footnote) to zero. The condition that k is small ensures that the policy associated with G^L is $d = 1 - \varepsilon$, which facilitates calculations. When k is large, the only difference is that the policy $d > d^*$ that accompnies G^L is given by $\partial U(G^L, d; \alpha)/\partial d = 0$.

leverage that it gives to the false narrative. In addition, α is hump-shaped with respect to δ. When μ is low, an improvement in μ via an increase in δ can make G^L *more* prevalent in equilibrium. The reason is that the narrative's ability to instill a false hope hinges on having enough historical variation in y. An increase in δ in its low region increases the variability of y, and therefore leaves more room for the easy-fix narrative to stoke false hope by attributing this variation in y to the wrong cause. Thus, demagogic narratives can actually become *more* popular when the underlying situation is better.

4 Analysis

The illustrative examples raise the question of whether policy divergence is an inherent feature of equilibrium. Our first result answers in the affirmative.

Proposition 2 *Suppose $C'' > 0$. Then, in any equilibrium (α, σ), σ assigns positive probability to exactly two policies, $d_r \geq d^*$ and $d_l \leq d^*$.*

Thus, the support of the equilibrium policy distribution consists of two elements that lie (weakly) on different sides of d^*. Unlike other results in this paper, the proof does not make heavy use of the DAG formalism. Rather, it relies heavily on non-nullness and the NSQD constraint. The intuition for the result is as follows. NSQD implies that if the equilibrium distribution assigned probability one to a single policy d, prevailing narratives cannot distort the consequences of this policy. But then non-nullness implies that some other narrative-policy pair will generate higher anticipatory utility. For instance, if $d \neq d^*$, a "narrator" can promote the ideal policy d^* with a "denialist" narrative that a has no effect on y. NSQD is also instrumental in establishing that there cannot be more than two policies in equilibrium, because it implies piecewise-linearity (with respect to d) of the indirect gross anticipatory utility.

The equilibrium policy distribution is a "mixture over mixtures": It is not the usual case of "mixed" equilibrium. Our assumption that policies are mixtures that generate *non-linear* intrinsic utility was meant to drive this point home. The fundamental insight behind this effect is that narratives can convey hopeful illusions only when coupled with *counterfactual* policies.

Remark 1 *Proposition 2 allows d_r or d_l to coincide with d^*. Slight modifications of non-nullness rule out this possibility. For instance, suppose that $\mathcal{G}$ includes two DAGs G and G' such that $p_G(y \mid a)$ and $p_{G'}(y \mid a)$ are strictly increasing and strictly decreasing in a, respectively. Then, $d_l < d^* < d_r$.*

4.1 Short Narratives

In this sub-section we characterize equilibria when narrators can use at most one variable in addition to a and y (i.e., $n = 3$). We focus on the case in which a and y are objectively *independent*. In this setting, the only narratives that can generate a non-constant $p_G(y \mid a)$ are the lever and threat/opportunity narratives. Our objective is to examine which of the two narratives will prevail, and which auxiliary variables they will employ. For this purpose, we assume that $m \gg n$ and the supply of potential auxiliary variables is rich, such that narrators can select the third variable in their narrative from an "ocean" of potential variables.

To introduce our notion of richness, let z be an arbitrary binary variable, and define Q^* to be the set of all conditional distributions $(p(z \mid a, y))$ such that $p(z \mid a)p(z \mid y) = 0$ for some values of a, y, z. That is, a conditional distribution in Q^* allows one value of a or one value of y to pin down deterministically the value of z.

Four particular elements in Q^* will play a special role. These are degenerate conditional distributions for which $p(z = 1 \mid a, y) \in \{0, 1\}$ for *every* a, y. Specifically, define

$$\begin{aligned}
q_1^{\wedge} &: \quad z = \mathbf{1}(a = 1 \text{ and } y = 1) \\
q_1^{\vee} &: \quad z = \mathbf{1}(a = 1 \text{ or } y = 1) \\
q_0^{\wedge} &: \quad z = \mathbf{1}(a = 0 \text{ and } y = 1) \\
q_1^{\vee} &: \quad z = \mathbf{1}(a = 0 \text{ or } y = 1)
\end{aligned}$$

We say that two sets of conditional distributions are close if the Hausdorff distance between any pair of elements from the two respective sets is below some arbitrarily small threshold.

Definition 3 *Let $m \gg n = 3$. An objective distribution p satisfying $a \perp y$ is* Q^*-rich *if $\{(p(x_i = 1 \mid a, y))\}_{i=2,\ldots,m-1}$ and Q^* are close.*

Q^*-richness says that the set of conditional distributions $(p(z \mid a, y))$ that one can simulate by selecting one auxiliary variable approximately coincides with Q^*. We impose this domain restriction because on one hand it is relatively weak (thus allowing for a large supply of potential auxiliary variables), yet on the other hand it is tractable. Numerical simulations suggest that the results of this sub-section will continue to hold if we replace Q^* with the set of *all* conditional distributions $(p(z \mid a, y))$.

Proposition 3 *For a generic Q^*-rich distribution p, there is an essentially unique equilibrium (α, σ):*[11]
(i) The policy $d_r > \alpha$ is accompanied by a lever narrative $a \to x_r \to y$, where $(p(x_r \mid a, y))$ is close to $q_1^{\wedge}$ or $q_1^{\vee}$.
(ii) The policy $d_l < \alpha$ is accompanied by a lever narrative $a \to x_l \to y$, where $(p(x_l \mid a, y))$ is close to $q_0^{\wedge}$ or $q_0^{\vee}$.

Thus, for generic Q^*-rich distributions, lever narratives prevail. The reason is that narratives that induce rational expectations generate lower anticipatory payoff, whereas threat/opportunity narratives violate NSQD. (The foreign policy example of Section 3.1 involved a non-generic distribution in this regard.) The lever narratives employ lever variables that are determined by two conditions: $y = 1$ and a takes the value that the narrative pushes for. The only remaining detail is whether these conditions are individually sufficient or jointly necessary. When C and ε are low, the resolution of this detail is simple.

[11] By essential uniqueness we mean that the policy distribution and the conditional distribution of the lever variables are unique (but the identity of the lever variables is not necessarily pinned down).

Remark 2 *Let C and ε be vanishingly small. Then, the equilibrium is as follows:*
(i) $\alpha = \frac{1}{2}$.
(ii) If $\mu < \frac{1}{2}$, *then* $(p(x_r \mid a, y)) \approx q_1^{\vee}$ *and* $(p(x_l \mid a, y)) \approx q_0^{\vee}$.
(ii) If $\mu > \frac{1}{2}$, *then* $(p(x_r \mid a, y)) \approx q_1^{\wedge}$ *and* $(p(x_l \mid a, y)) \approx q_0^{\wedge}$.

For real-life examples of lever narratives that are captured by this equilibrium characterization, recall the US trade policy debate described in the Introduction. In this context, our characterization approximates the following prevailing narratives. The lever narrative that sustains a policy with a protectionist bias (relative to the agent's ideal point) will involve a variable like "imports from China", because low imports are associated with trade restrictions as well as high employment in the local manufacturing sector. The narrative is false if the latter correlation is not causal but due to a confounding factor (such as exogenous technology changes that affect outsourcing of production). Likewise, the lever narrative that sustains a trade policy with a liberalized bias will select a variable like "industrial exports".

The argument that delivers equilibrium uniqueness in Proposition 3 is a "diminishing returns" property of the kind we discussed in Section 3.1. That is, the right-wing (left-wing) narrative's induced anticipatory utility decreases (increases) with α. This property also means that the unique equilibrium is dynamically stable: Perturbing α in one direction makes the opposite narrative more appealing, which gives rise to an equilibrating force.

5 State-Dependent Narrative Selection

So far, we have assumed that narrative-policy pairs are evaluated without conditioning on any variable. However, the appeal of a given narrative often varies with changing circumstances. In this section, we extend the definition of equilibrium in this direction and illustrate the extended concept.

Recall the collection of variables $x_1, ..., x_m$, where x_1 (also denoted a) is the action and x_m (also denoted y) is the consequence. Let $m \geq 3$ and assume that the variable x_2 (also denoted θ) is realized and publicly observed before the narrative-policy pair is evaluated. We refer to θ as a "state variable". Given a DAG G and an objective distribution p, the subjective belief p_G is defined as before, and the subjective conditional distribution $p_G(y \mid a, \theta)$ is defined as usual. Define the gross conditional anticipatory utility

$$V(G, d \mid \theta) = d \cdot p_R(y = 1 \mid \theta, a = 1) + (1 - d) \cdot p_R(y = 1 \mid \theta, a = 0)$$

The agent's net anticipatory utility is $U(G, d \mid \theta) = V(G, d \mid \theta) - C(d - d^*)$.[12] For every θ, define $\alpha_\theta = p(a = 1 \mid \theta)$, and let σ_θ denote a distribution over narrative-policy pairs (G, d) given θ. Denote $\alpha = (\alpha_\theta)_\theta$ and $\sigma = (\sigma_\theta)_\theta$.

Definition 4 *A pair (α, σ) is an equilibrium if, for every θ:*
(i) For every $(G, d) \in Supp(\sigma_\theta)$, $(G, d) \in \arg\max_{(G,d)} U(G, d \mid \theta)$ subject to the constraint that $\sum\limits_\theta p(\theta)V(G, d \mid \theta) = \mu$.
(ii) $\alpha_\theta = \sum_{(G,d)} \sigma_\theta(G, d) \cdot d$

At first glance, this may seem like an uninteresting state-by-state extension of our equilibrium concept. However, note that the NSQD constraint is global. In addition, depending on how a narrative treats θ, the objective distribution of certain variables at one value of θ can influence their subjective distribution conditional on another value. This externality is what makes this extension interesting, as the following example demonstrates.

An example: Denialism and exaggeration
Let $m = n = 3$, where the three variables are the action a, the consequence y and the state variable θ. Let $p(\theta = 1) = \delta$ and $p(y = 1 \mid a, \theta) = \frac{1}{2}(a + \theta)$. Let $d^* = \varepsilon$, where ε is arbitrarily small. Assume C is continuous, twice-differentiable and steep enough such that $C'(1) > 1$. Because $p(y = 1 \mid a, \theta)$ is additively separable, the optimal policy under rational expectations is $d^{RE} = \arg\max_d(\frac{1}{2}d - C(d))$, independently of θ. By the assumptions on C, d^{RE} is interior and given by $C'(d^{RE}) = \frac{1}{2}$.

[12] Here (unlike the rest of the paper) we suppress α in V and U, in order to keep the notation non-fussy.

Let $\mathcal{G}$ be the set of all DAGs in which y is a terminal node and there is a direct link $\theta \to a$. The interpretation is that the representative agent is aware that actions are taken in response to θ; consequently, any narrative should incorporate this manifest causal relation. Then, $\mathcal{G}$ consists of the following four DAGs: $G^d : a \leftarrow \theta \to y$; $G^e : \theta \to a \to y$; the DAG G^n that removes the link $a \to y$ from G^e; and the fully connected DAG G^{RE} that adds the link $\theta \to y$ to G^e. The latter is the only DAG in $\mathcal{G}$ that is consistent with p, because all the others rule out the direct effect of a or θ on y. All DAGs in $\mathcal{G}$ are perfect, hence we can take NSQD for granted.

One concrete story for this example involves debates over exploitation of scarce natural resources. In this context, θ represents exogenous fluctuations in the availability of this resource, y represents the resource's net availability, and a represents preservation policy, where $a = 1$ stands for costly preservation measures (hence the assumption that $d^* = \varepsilon$). Accordingly, G^d is a "*denialist*" narrative that neglects the role of policy and attributes the consequence entirely to exogenous forces. In contrast, G^e is an "*exaggerationist*" narrative that effectively says "it is all up to us". The DAG G^n is a "*neutral*" narrative because it does not attribute the outcome to any of the other variables.

Claim 3 *There is a unique equilibrium, which is characterized as follows:*
(i) $Supp(\sigma_1) = \{(G^{RE}, d^{RE}), (G^d, \varepsilon)\}$.
(ii) $Supp(\sigma_0) = \{(G^e, d^{RE}), (G^n, \varepsilon)\}$.
(iii) $\alpha_1 = \alpha_0 = d^{RE} - 2C(d^{RE})$.

Thus, different states give rise to the same mixture between policies, but these policies are promoted by *different sets of narratives.* The rational and denialist narratives prevail in the good state $\theta = 1$, whereas the exaggerationist and neutral narratives prevail in the bad state $\theta = 0$. In each case, narratives that neglect the role of a legitimize the representative agent's desire to eschew hard trade-offs, and therefore induce the ideal policy $d^* = \varepsilon$. And the narratives that account for the role of a induce the rational-expectations policy, even if they do not always give it the rational-expectations rationale.

The result that the mixture over policies is state-independent mirrors the rational-expectations benchmark. However, the reasoning behind it is subtler. The narratives G^e and G^n effectively fail to condition anticipatory utility on θ. As a result, there is an externality between the two states that does not exist under rational expectations. In particular, α_1 affects the relative appeal of G^e and G^n in $\theta = 1$, and therefore could potentially affect α_0. It is *equilibrium reasoning* that restores state-independent policy mixtures: If α_1 were higher (lower) than α_0, this would make G^e more (less) appealing, thus leading to a rise (drop) in α_0.

6 Conclusion

The model of competing narratives presented in this paper formalized a number of intuitions regarding the role of narratives in the formation of political beliefs. Our model was based on two main ideas.

What are narratives and how do they shape beliefs? In our formalism, narratives are causal models that map actions into consequences. Different narratives employ different intermediate variables and arrange them differently in the causal scheme. Narratives shape beliefs by imposing a causal interpretation on long-run correlations. These beliefs are used to evaluate policies.

How does the public select between competing narratives? Our behavioral assumption was that in the presence of conflicting narrative-policy pairs, the public (a representative agent in this paper) selects between them according to their induced anticipatory utility. This is consistent with the basic intuition that people are drawn to stories with a "hopeful" message.

The main insights that emerged from our analysis of the model can be summarized as follows. First, some prevailing narratives are misspecified causal models that "sell false hopes" regarding the consequences of counterfactual policies. Second, the same variable can serve two conflicting narratives with a different causal structure in the service of conflicting policies. Third, multiplicity of dominant narrative-policy pairs is an intrinsic property of long-run equilibrium in the "battle over public opinion". Indeed, in specific settings, we saw that growing popularity of one policy weakens the appeal of its supporting narrative. This "diminishing returns" aspect leads to additional properties of equilibrium (uniqueness, dynamic stability) in these settings. Finally, we hope that our stylized examples gave a foretaste of the model's ability to explain the popularity of certain real-life political narratives and their implications for political outcomes.

References

[1] Akerlof, G. and W. Dickens (1982), The economic consequences of cognitive dissonance, *American Economic Review* 72, 307-319.

[2] Bénabou, R. and J. Tirole (2002), Self-Confidence and Personal Motivation, *Quarterly Journal of Economics* 117, 871–915.

[3] Benabou, R. and J. Tirole (2016), Mindful Economics: The Production, Consumption and Value of Beliefs, *Journal of Economic Perspectives* 30, 141-164.

[4] Benabou, R., A. Falk and J. Tirole (2018), Narratives, Imperatives and Moral Reasoning, NBER Working Paper No. 24798.

[5] Brunnermeier, M. and J. Parker (2005), Optimal Expectations, *American Economic Review* 95, 1092-1118.

[6] Caron, R. and T. Traynor (2005), The Zero Set of a Polynomial, WSMR Report: 05-02.

[7] Cowell, R., P. Dawid, S. Lauritzen and D. Spiegelhalter (1999), *Probabilistic Networks and Expert Systems,* Springer, London.

[8] Dahlstrom, M. F. (2014), Using narratives and storytelling to communicate science with nonexpert audiences, *Proceedings of the National Academy of Science* 111, 13614–13620.

[9] De Graaf, B., G. Dimitriu and J. Ringsmose (2016), *Strategic Narratives, Public Opinion and War: Winning domestic support for the Afghan War,* Taylor and Francis.

[10] Esponda, I. and D. Pouzo (2017), Retrospective Voting and Party Polarization, *International Economic Review*, forthcoming.

[11] Frum, D. (2019), If Liberals Won't Enforce Borders, Fascists Will, The Atlantic Magazine, https://www.theatlantic.com/magazine/archive/2019/04/david-frum-how-much-immigration-is-too-much/583252/.

[12] Levy, G. and R. Razin (2018), An Explanation-Based Approach to Combining Forecasts, mimeo.

[13] McComas, K. and J. Shanahan (1999), Telling Stories About Global Climate Change: Measuring the Impact of Narratives on Issue Cycles, *Communication Research* 26, 30-57.

[14] Monteal Olea, J., P. Ortoleva, M. Pai and A. Prat (2018), Competing Models, mimeo.

[15] Pearl, J. (2009), *Causality: Models, Reasoning and Inference,* Cambridge University Press, Cambridge.

[16] Schwartzstein, J. and A. Sunderam (2019), Using Models to Persuade, mimeo.

[17] Shiller, R. (2017), Narrative Economics, *American Economic Review* 107, 967-1004.

[18] Sloman, S. (2005), *Causal Models: How People Think about the World and its Alternatives*, Oxford University Press.

[19] Sloman, S. and D. Lagnado (2015), Causality in Thought, *Annual Review of Psychology* 66, 223-247.

[20] Spiegler, R. (2008), On Two Points of View Regarding Revealed Preferences and Behavioral Economics, in *The Foundations of Positive and Normative Economics*, Oxford University Press, 95-115.

[21] Spiegler, R. (2013), Placebo Reforms, *American Economic Review* 103, 1490-1506.

[22] Spiegler, R. (2016), Bayesian Networks and Boundedly Rational Expectations, *Quarterly Journal of Economics* 131, 1243-1290.

[23] Spiegler, R. (2018), Can Agents with Causal Misperceptions be Systematically Fooled? *Journal of the European Economic Association*, forthcoming.

[24] Szostek, J. (2018). Nothing Is True? The Credibility of News and Conflicting Narratives during 'Information War' in Ukraine. *The International Journal of Press/Politics* 23, 116-135.

Appendix I: Proofs

Proof of Proposition 1

Consider an auxiliary two-player game. Player 1's strategy space is D, and α denotes an element in this space. Player 2's strategy space is $\Delta(\mathcal{G} \times D)$, and β denotes an element in this space. The payoff of player 1 from the strategy profile (α, β) is equal to $\sum_{G,d} \beta(G,d)\widetilde{U}(G,d;\alpha)$, where $\widetilde{U}(G,d;\alpha) = U(G,d;\alpha)$ if $V(G,\alpha;\alpha) = \mu$ and $\widetilde{U}(G,d;\alpha) = -1$ otherwise. The payoff of player 2 from (σ, α) is $-(\alpha - \sum_{G,d} \beta(G,d)d)^2$. A Nash equilibrium in this auxiliary game is equivalent to our notion of equilibrium. Hence, our objective is to establish existence of a Nash equilibrium (α, β) in which every narrative in the support of β satisfies NSQD with respect to α. Since p_G is a continuous function of α, so is U. In addition, the strategy spaces and payoff functions of the two players in the auxiliary game satisfy standard conditions for the existence of Nash equilibrium. Non-nullness ensures that $\mathcal{G}$ includes a DAG G^* that induces $V(G,\alpha;\alpha) = \mu$. This means that (G^*, α) strictly dominates any (G,d) with $V(G,\alpha;\alpha) \neq \alpha$ in the auxiliary game. It follows that the support of β in any Nash equilibrium does not include pairs (G,d) such that G violates NSQD. ■

Proof of Claim 2

The restrictions on $\mathcal{G}$ and p satisfy the conditions of Proposition 2. Therefore, σ assigns positive probability to exactly two policies coupled to different narratives. Thus, $Supp(\sigma) = \{(G^*, d^1), (G^L, d^2)\}$. Recall that $p(s=1|a,\theta) = (1-\rho)(a+(1-a)\theta)$. The analysis that follows takes ρ to be vanishingly small.

Consider G^*. Clearly, $(p_{G^*}(y \mid a)) = (p(y \mid a))$. Since a and y are objectively independent, it follows that the policy that maximizes net anticipatory utility under p_{G^*} is d^*. Therefore, the policy that accompanies G^* is $d^1 = d^*$.

Now consider G^L:

$$p_{G^L}(y=1 \mid a) = p(s=1 \mid a)p(y=1 \mid s=1) + p(s=0 \mid a)p(y=1 \mid s=0)$$

Note that $p(s=1 \mid a=1) = 1$, such that

$$p_{G^L}(y=1 \mid a=1) = p(y=1 \mid s=1) = \frac{\delta\beta + (1-\delta)(1-\beta)\alpha}{\delta + (1-\delta)\alpha}$$

To calculate $p_{G^L}(y=1 \mid a=0)$, we note that $p(y=1 \mid s=0) = 1-\beta$ (because $s=0$ can only happen when $\theta=0$, in which case $y=1$ with probability $1-\beta$). Also, $p(s=1 \mid a=0) = \delta$, because if $a=0$ then $s=1$ if and only if $\theta=1$. Then,

$$p_{G^L}(y=1 \mid a=0) = (1-\delta)(1-\beta) + \delta \cdot \frac{\delta\beta + (1-\delta)(1-\beta)\alpha}{\delta + (1-\delta)\alpha}$$

Therefore,

$$p_{G^L}(y=1 \mid a=1) - p_{G^L}(y=1 \mid a=0) = \frac{\delta(1-\delta)(2\beta-1)}{\delta + (1-\delta)\alpha}$$

which is strictly positive because $\beta > \frac{1}{2}$. Since the derivative of $U(G^L, d; \alpha)$

with respect to d is

$$U'(G^L, d; \alpha) = p_{G^L}(y = 1 \mid a = 1) - p_{G^L}(y = 1 \mid a = 0) - C'(d - d^*)$$

it follows that $d^2 > d^*$. Furthermore, since $C'(d - d^*) = 2k(d - d^*)$, $U'(G^L, d; \alpha) > 0$ for every d, thanks to the restriction on k. Therefore, $d^2 = 1 - \varepsilon$. To confirm that this is consistent with equilibrium, we need to verify that $U(G^L, d^2; \alpha) = U(G^*, d^*; \alpha)$. For d^2 sufficiently close to 1, we have that

$$\frac{\delta\beta + (1-\delta)(1-\beta)\alpha}{\delta + (1-\delta)\alpha} - k(1 - d^*)^2 \approx \beta\delta + (1-\beta)(1-\delta)$$

Solving this equation gives the solution for α, which is strictly between 0 and 1 thanks to the restriction on k. ■

Proof of Proposition 2

Fix an equilibrium (α, σ). First, let us show that every $(G, d) \in Supp(\sigma)$ induces $U(G, d; \alpha) \geq \mu$. Assume the contrary. By minimal richness, $\mathcal{G}$ includes the DAG $G^* : a \quad y$. Note that $p_G(y \mid a) = \mu$ for every a. It follows that the narrative-policy pair (G^*, d^*) generates the net payoff $U(G^*, d^*; \alpha) = \mu$, contradicting the first part of the definition of equilibrium.

Next, we establish that the support of σ must include at least two distinct policies. Assume the contrary - i.e., the marginal of σ over d is degenerate. Then by definition, it assigns probability one to the steady-state policy α. By NSQD, $V(G, \alpha \mid \alpha) = \mu$ for every narrative G in the support of σ.

There are two cases to consider. First, suppose $\alpha \neq d^*$. Then, any narrative G in the support of σ delivers $U(G, \alpha; \alpha) = \mu - C(\alpha - d^*)$. By assumption, $C'' > 0$ such that $C(\alpha - d^*) > 0$. Therefore, $U(G, \alpha; \alpha) < \mu$, contradicting our previous step. Second, suppose $\alpha = d^*$. Then, any narrative G in the support of σ delivers $U(G, \alpha; \alpha) = \mu$. By minimal richness, $\mathcal{G}$ contains the DAG $G^{**} : a \to x_i \to y$, where x_i is correlated with both a and y according to p. Without loss of generality, suppose $p(x_i = 1 \mid a = 1) > p(x_i = 1 \mid a = 0)$ and $p(x_i = 1 \mid a = 1) > p(x_i = 1 \mid a = 0)$. Since G^{**} is perfect, it satisfies NSQD, such that $U(G^{**}, \alpha; \alpha) = \mu$. The derivative of $V(G^{**}, d; \alpha)$ with respect to d is $p_{G^{**}}(y = 1 \mid a = 1) - p_{G^{**}}(y = 1 \mid a = 0)$, which can be written as

$$[p(x_i = 1 \mid a = 1) - p(x_i = 1 \mid a = 0)]\,[p(y = 1 \mid x_i = 1) - p(y = 1 \mid x_i = 0)]$$

By assumption, both terms in this product are non-zero, hence the derivative of $V(G^{**}, d; \alpha)$ with respect to d is non-zero. Since the derivative of C at $d = d^*$ is zero, it follows that there is $d \neq d^*$ such that $U(G^{**}, d; \alpha) > \mu$, again contradicting the first part in the definition of equilibrium.

We now show that the support cannot contain more than two policies. By definition, any $(G, d) \in Supp(\sigma)$ maximizes $V(G, d; \alpha) - C(d - d^*)$ subject to the NSQD constraint $V(G, \alpha; \alpha) = \mu$. This means that we can rewrite $V(G, d; \alpha)$ as follows:

$$\begin{aligned} V(G, d; \alpha) &= \frac{d-\alpha}{1-\alpha} \cdot p_G(y = 1 \mid a = 1) + \frac{1-d}{1-\alpha} \cdot \mu \\ &= (1 - \frac{d}{\alpha}) \cdot p_G(y = 1 \mid a = 0) + \frac{d}{\alpha} \cdot \mu \end{aligned} \tag{9}$$

It follows that the narratives that maximize U for a given (d, α) subject to NSQD depend only on the *ordinal* ranking between d and α. That is, for every $(G, d) \in Supp(\sigma)$ such that $d \geq \alpha$, G maximizes $p_G(y = 1 \mid a = 1)$. Likewise, for every $(G, d) \in Supp(\sigma)$ such that $d \leq \alpha$, G maximizes $p_G(y = 1 \mid a = 0)$. Note that when $d = \alpha$, any G that satisfies NSQD maximizes U.

This means that $\max_G V(G, d; \alpha)$ is piecewise linear in d, having a weakly positive slope in the range $d \geq \alpha$ and a weakly negative slope in the range $d \leq \alpha$, where at least one of these slopes is non-zero. Because C is strictly convex, it follows that there exist unique d_r and d_l that maximize U in the ranges $d \geq \alpha$ and $d \leq \alpha$, respectively. It follows that there are at most two policies in the support of the equilibrium distribution, and that they lie (weakly) on different sides of α.

It remains to establish that $d_r \geq d^*$ and $d_l \leq d^*$. We have already shown that it cannot be the case that $d_r = d_l = d^*$. Assume d_r and d_l are (strictly) on the same side of d^*. Without loss of generality, let $d_r > d_l > d^*$. The second part of the definition of equilibrium implies $d_l < \alpha < d_r$. Since $d_l < \alpha$, we saw that if $(G, d_l) \in Supp(\sigma)$, then G maximizes $p_G(y = 1 \mid a = 0)$. Since

$d^* < d_l$, it follows that the pair (G, d^*) would attain a strictly higher U than (G, d_l), contradicting the first part of the definition of equilibrium.

Proof of Proposition 3

The assumption that p is Q^*-rich enables us to apply Remark 1: In any equilibrium, the support of the marginal equilibrium distribution over policies is $\{d_l, d_r\}$, where $d_l < d^* < d_r$. Furthermore, the policies d_r and d_l are accompanied by narratives G_r and G_l that maximize $p_G(y = 1 \mid a = 1)$ and $p_G(y = 1 \mid a = 0)$, respectively.

The proof proceeds stepwise.

Step 1: For any $k = 2, ..., m-1$, the threat/opportunity DAG $1 \rightarrow m \leftarrow k$ violates NSQD for almost all rich distributions p.

Proof: Let $G : a \rightarrow y \leftarrow z$, where $z \in \{0, 1\}$ and $(p(z \mid a, y))$ is a generic element in Q^*. Since a is an ancestral node in G, we can substitute $p(a) \equiv p_G(a)$ (see Spiegler (2017)), such that the NSQD requirement can be written as

$$\sum_a p_G(a) p_G(y = 1 \mid a) = p(y = 1) = \mu$$

Since the L.H.S of this equation is by definition $p_G(y = 1)$, it follows that NSQD is equivalent to the requirement that p_G does not distort the objective marginal distribution of y. We can write the condition more explicitly:

$$\sum_a \sum_z p(a) \left(\sum_{a'} \sum_{y'} p(a') p(y') p(z \mid a', y') \right) \frac{p(a) \mu p(z \mid a, y = 1)}{p(a) \sum_{y''} p(y'') p(z \mid a, y'')} = \mu$$

This expression can be simplified into

$$\sum_a p(a) \sum_z \frac{p(z \mid a, y = 1) \sum_{a'} \sum_{y'} p(a') p(y') p(z \mid a', y')}{\sum_{y''} p(y'') p(z \mid a, y'')} = 1$$

This is an equation in four variables $(p(z = 1 \mid a, y))$, where $(p(a))$ and $(p(y))$ are given constants. We can multiply both sides of the equations by the four terms $(\sum_{y''} p(y'') p(z \mid a, y''))_{a,z}$, and obtain a polynomial equation in the four variables. The equation is non-tautological: it is violated when

$z \approx y + a(1-y)$.[13] It is well-known that the Lebesgue measure of the set of solutions of a non-tautological polynomial equation over $[0,1]^n$ is zero (see Caron and Traynor (2005)). This completes the proof. □

Thus, for generic Q^*-rich distributions p, the only DAGs G that can be part of an equilibrium while inducing non-constant $p_G(y \mid a)$ are the lever DAGs $a \to x_i \to y$, where $i = 2, ..., m-1$. The narratives that accompany d_r and d_l both have this structure, and thus only differ in the value of i. Q^*-richness means that the problem of finding the value of i for G_r is approximated by the following problem:[14]

$$\max_{(p(z=1|a,y))_{a,y=0,1} \in Q^*} \sum_{z=0,1} p(z \mid a=1) p(y=1 \mid z) \tag{10}$$

$$= \sum_z \left(\sum_{y'} p(y') p(z \mid a=1, y') \right) \frac{\mu \sum_{a'} p(a') p(z \mid a', y=1)}{\sum_{y''} \sum_{a''} p(a'') p(y'') p(z \mid a'', y'')}$$

The problem for G_l is the same, except that we condition on $a = 0$ instead of $a = 1$.

Step 2: The solution to (10) is

$$p_G(y=1 \mid a) = \max \left\{ \frac{\mu}{\mu + p(a)(1-\mu)}, \frac{\mu(2 - p(a) - \mu)}{1 - \mu p(a)} \right\}$$

where the left and right arguments are attained at $q_a^\vee$ and $q_a^\wedge$, respectively. The left argument is weakly higher than the right argument if and only if $p(a) + \mu \leq 1$. Denote the solution by $H_a(\alpha)$.

Proof: See Appendix II. □

Step 3: The equilibrium is generically unique.

[13] Suppose z is determined as follows: With probability $1-\rho$, $z = y + a(1-y)$, and with probability ρ, $z = 0$. For ρ sufficiently close to zero,

$$\sum_a p_G(a) p_G(y=1 \mid a) \approx [1 - \alpha(1-\mu)][\alpha + (1-\alpha)\mu] > \mu$$

[14] This is only an approximation because we would need to incorporate small perturbations to the distribution of the lever variable if the solution to the maximization problem yields distributions without full support for every realization of a and y.

Proof: By Q^*-richness, $p_{G_r}(y = 1 \mid a = 1) \approx H_1(\alpha)$ and $p_{G_l}(y = 1 \mid a = 0) \approx H_0(\alpha)$. Use NSQD to define $p_{G_r}(y = 1 \mid a = 0)$ and $p_{G_l}(y = 1 \mid a = 1)$ in terms of $H_a(\alpha)$, and obtain

$$
\begin{aligned}
V(G_r, d; \alpha) &\approx \mu \cdot \max\left\{ \frac{d(1-\mu)+\mu}{\alpha(1-\mu)+\mu}, \frac{d(1-\mu)+1-\alpha}{\alpha(1-\mu)+1-\alpha} \right\} \\
V(G_l, d; \alpha) &\approx \mu \cdot \max\left\{ \frac{(1-d)(1-\mu)+\mu}{(1-\alpha)(1-\mu)+\mu}, \frac{(1-d)(1-\mu)+\alpha}{(1-\alpha)(1-\mu)+\alpha} \right\}
\end{aligned}
$$

Consider an equilibrium with some given α. By definition,

$$
\begin{aligned}
U(G_r, d_r; \alpha) &= \max_{d>\alpha}[V(G_r, d; \alpha) - C(d - d^*)] \\
U(G_l, d_l; \alpha) &= \max_{d<\alpha}[V(G_l, d; \alpha) - C(d - d^*)]
\end{aligned}
$$

It is easy to verify that for any fixed d, $V(G_r, d; \alpha)$ is strictly decreasing with α, whereas $V(G_l, d; \alpha)$ is strictly increasing with α. Consequently, the equation $U(G_r, d_r; \alpha) = U(G_l, d_l; \alpha)$ that must hold in equilibrium cannot have more than one solution α. Given α, the solution of G_r and G_l is generically unique, and therefore $V(G_r, d; \alpha)$ and $V(G_l, d; \alpha)$ are also pinned down. As a result, d_r and d_l are pinned down, which also pins down σ.[15]

[15] Here, genericity means that when $p(a) + \mu = 1$, there are two lever variables that maximize $p_G(y = 1|a = 1)$, $z = ay$ and $z = y + a(1 - y)$; and two lever variables that maximize $p_G(y = 1|a = 0)$, $x = y(1 - a)$ and $x = y + (1 - y)(1 - a)$. For details, see Appendix 2.

Proof of Remark 2

When C and ε are vanishingly low, it must be the case that $d_r \approx 1$ and $d_l \approx 0$, such that

$$\max\left\{\frac{2-\mu-\alpha}{\alpha(1-\mu)+1-\alpha}, \frac{1}{\alpha(1-\mu)+\mu}\right\} \approx \max\left\{\frac{1-\mu+\alpha}{(1-\alpha)(1-\mu)+\alpha}, \frac{1}{(1-\alpha)(1-\mu)+\mu}\right\}$$

The result follows by solving this equation.

Proof of Claim 3

Let us calculate $p_G(y = 1 \mid a, \theta)$ for each of the four available narratives:

$$\begin{aligned}
p_{G^{RE}}(y &= 1 \mid a, \theta) = p(y = 1 \mid a, \theta) = \frac{1}{2}(a + \theta) \\
p_{G^n}(y &= 1 \mid a, \theta) = p(y = 1) = \frac{1}{2}[\delta(1 + \alpha_1) + (1 - \delta)\alpha_0] \\
p_{G^d}(y &= 1 \mid a, \theta) = p(y = 1 \mid \theta) = \frac{1}{2}(\alpha_\theta + \theta) \\
p_{G^e}(y &= 1 \mid a, \theta) = p(y = 1 \mid a) = \frac{1}{2}[a + p(\theta = 1 \mid a)]
\end{aligned}$$

where

$$\begin{aligned}
p(\theta &= 1 \mid a = 1) = \frac{\delta\alpha_1}{\delta\alpha_1 + (1 - \delta)\alpha_0} \\
p(\theta &= 1 \mid a = 0) = \frac{\delta(1 - \alpha_1)}{\delta(1 - \alpha_1) + (1 - \delta)(1 - \alpha_0)}
\end{aligned}$$

It follows that the net anticipatory utility induced by a policy d coupled with any of the four narratives is:

$$\begin{aligned}
U(G^{RE}, d \mid \theta) &= \frac{1}{2}\theta + \frac{1}{2}d - C(d) \\
U(G^n, d \mid \theta) &= \frac{1}{2}[\delta(1 + \alpha_1) + (1 - \delta)\alpha_0] - C(d) \\
U(G^d, d \mid \theta) &= \frac{1}{2}(\alpha_\theta + \theta) - C(d) \\
U(G^e, d \mid \theta) &= \frac{1}{2}d - C(d) + \frac{1}{2}\left[\frac{\delta\alpha_1 d}{\delta\alpha_1 + (1 - \delta)\alpha_0} + \frac{\delta(1 - \alpha_1)(1 - d)}{\delta(1 - \alpha_1) + (1 - \delta)(1 - \alpha_0)}\right]
\end{aligned}$$

The policy that maximizes net anticipatory utility under G^d or G^n is

$d^* = 0$. Therefore, if any of these narratives prevails in some state, it must be coupled with $d = 0$. Likewise, the policy that maximizes net anticipatory utility under G^{RE} is d^{RE}. Therefore, if this narrative prevails in some state, it must be coupled with d^{RE}. As to the narrative G^e, the policy d^e that maximizes net anticipatory utility under this narrative satisfies $d^e > d^{RE}$ ($d^e < d^{RE}$) whenever $\alpha_1 > \alpha_0$ ($\alpha_1 < \alpha_0$).

Note that it follows from $C'(1) > 1$ that even under the most optimistic belief that is induced by one of the narratives, the optimal policy would always be strictly below 1. Hence, $\alpha_\theta < 1$ for all θ.

Consider the realization $\theta = 1$. Suppose $\alpha_1 = 0$. Then,

$$U(G^{RE}, d^{RE} \mid \theta = 1) = \frac{1}{2} + \max_d [\frac{1}{2}d - C(d)] > \frac{1}{2}$$

whereas

$$\begin{aligned} U(G^n, 0 \quad | \quad \theta = 1) &= \frac{1}{2}[\delta + (1-\delta)\alpha_0] < \frac{1}{2} \\ U(G^d, 0 \quad | \quad \theta = 1) &= \frac{1}{2} \end{aligned}$$

In addition, for any d and for all α_0,

$$\frac{1}{2}d \cdot \frac{(1-\delta)(1-\alpha_0)}{\delta + (1-\delta)(1-\alpha_0)} - C(d) + \frac{1}{2} \cdot \frac{\delta}{\delta + (1-\delta)(1-\alpha_0)} < \frac{1}{2}d - C(d) + \frac{1}{2}$$

which implies that $U(G^e, d^e \mid \theta = 1) < U(G^{RE}, d^{RE} \mid \theta = 1)$. Therefore, (G^{RE}, d^{RE}) must be the prevailing narrative-policy pair, contradicting the assumption that $\alpha_1 = 0$.

It follows that $\alpha_1 > 0$. Since for any α_0 and for any d,

$$\frac{\delta\alpha_1 d}{\delta\alpha_1 + (1-\delta)\alpha_0} + \frac{\delta(1-\alpha_1)(1-d)}{\delta(1-\alpha_1) + (1-\delta)(1-\alpha_0)} < 1 \tag{11}$$

we have that $U(G^e, d \mid \theta = 1) < U(G^{RE}, d \mid \theta = 1)$, and hence, G^e cannot be a prevailing narrative in $\theta = 1$. Likewise, a simple calculation establishes that $U(G^d, 0 \mid \theta = 1) > U(G^n, 0 \mid \theta = 1)$. Therefore, G^n is not a prevailing narrative in $\theta = 1$.

It follows that the only narrative-policy pairs that can prevail in $\theta = 1$ are (G^{RE}, d^{RE}) and $(G^d, 0)$. Their induced net anticipatory utility is

$$U(G^{RE}, d^{RE} \mid \theta = 1) = \frac{1}{2}(d^{RE} + 1) - C(d^{RE})$$
$$U(G^d, 0 \mid \theta = 1) = \frac{1}{2}(\alpha_1 + 1)$$

If $Supp(\sigma_1) = \{(G^d, 0)\}$, then $\alpha_1 = 0$, which we have already ruled out. Suppose $Supp(\sigma_1) = \{(G^{RE}, d^{RE})\}$. Then, $\alpha_1 = d^{RE}$, in which case it is obvious that $U(G^{RE}, d^{RE} \mid \theta = 1) < U(G^d, 0 \mid \theta = 1)$, a contradiction. The only remaining case is that $Supp(\sigma_1) = \{(G^d, 0), (G^{RE}, d^{RE})\}$. Then, $U(G^{RE}, d^{RE} \mid \theta = 1) = U(G^d, 0 \mid \theta = 1)$, which implies $\alpha_1 = d^{RE} - 2C(d^{RE})$. The first-order-condition characterization of d^{RE} and the strict convexity of C ensure that indeed, $\alpha_1 \in (0, 1)$. This completes the characterization of σ_1. Note that it is independent of σ_0.

Next, consider the realization $\theta = 0$. For any d, the difference, $U(G^e, d \mid \theta = 0) - U(G^{RE}, d \mid \theta = 0)$, is equal to half of the L.H.S. of (11), which is positive since $\alpha_1 > 0$. Therefore, G^{RE} cannot be a prevailing narrative in $\theta = 0$. Likewise, $U(G^n, 0 \mid \theta = 0) > U(G^d, 0 \mid \theta = 0)$, and hence, G^d cannot be a prevailing narrative in $\theta = 0$. It follows that in $\theta = 0$ the only narrative-policy pairs that can prevail are (G^e, d^e) and $(G^n, 0)$, where $d^e = \arg\max_d U(G^e, d \mid \theta)$ (from the strict convexity of C, this function has a unique maximum).

Let us guess an equilibrium in which $\alpha_0 = \alpha_1$. Then $U(G^e, d \mid \theta = 0) = \frac{1}{2}d - C(d) + \frac{1}{2}\delta$, and the policy that maximizes it is $d^e = d^{RE}$. Thus,

$$U(G^e, d^e \mid \theta = 0) = \frac{1}{2}d^{RE} - C(d^{RE}) + \frac{1}{2}\delta = \frac{1}{2}\alpha_1 + \frac{1}{2}\delta$$
$$U(G^n, 0 \mid \theta = 0) = \frac{1}{2}[\delta(1 + \alpha_1) + (1 - \delta)\alpha_1] = \frac{1}{2}\alpha_1 + \frac{1}{2}\delta$$

which is consistent with $\alpha_0 \in (0, 1)$.

We next show that there exists no equilibrium with $\alpha_0 \neq \alpha_1$. Suppose first that $\alpha_1 > \alpha_0$. Note that

$$\begin{aligned}\max_d U(G^e, d \quad | \quad \theta = 0) &\geq U(G^e, \alpha_1 \mid \theta = 0) \\ &= \frac{1}{2}\alpha_1 + \frac{1}{2}\delta\left[\frac{\alpha_1^2}{\delta\alpha_1 + (1-\delta)\alpha_0} + \frac{(1-\alpha_1)^2}{1-\delta\alpha_1-(1-\delta)\alpha_0}\right]\end{aligned}$$

We argue that if $\alpha_0 \neq \alpha_1$ then

$$\frac{\alpha_1^2}{\delta\alpha_1 + (1-\delta)\alpha_0} + \frac{(1-\alpha_1)^2}{1-\delta\alpha_1-(1-\delta)\alpha_0} > 1$$

A bit of algebra confirms that this inequality is satisfied if and only if $(\alpha_1 - \alpha_0)^2 > 0$. Hence, when $\alpha_0 \neq \alpha_1$,

$$\max_d U(G^e, d \mid \theta = 0) \geq U(G^e, \alpha_1 \mid \theta = 0) > \frac{1}{2}\alpha_1 + \frac{1}{2}\delta$$

But when $\alpha_0 < \alpha_1$,

$$U(G^n, 0 \mid \theta = 0) < \frac{1}{2}[\delta(1+\alpha_1) + (1-\delta)\alpha_1] = \frac{1}{2}\alpha_1 + \frac{1}{2}\delta$$

which implies that $Supp(\sigma_0) = \{(G^e, d^e)\}$, and hence, $\alpha_0 = d^e$. But when $\alpha_1 > \alpha_0$ we know that $d^e > d^{RE} = \alpha_1$. This implies that $\alpha_0 > \alpha_1$, a contradiction.

Suppose instead that $\alpha_0 > \alpha_1$. If $Supp(\sigma_0) = \{(G^e, d^e)\}$, then $\alpha_0 = d^e < d^{RE} = \alpha_1$, a contradiction. If $Supp(\sigma_0) = \{(G^n, 0)\}$, then $\alpha_0 = 0 < \alpha_1$, a contradiction. If $Supp(\sigma_0) = \{(G^e, d^e), (G^n, 0)\}$, then α_0 will be a convex combination of $d^e < \alpha_1$ and $d^n = 0$, which is strictly lower than α_1. But this contradicts our assumption that $\alpha_0 > \alpha_1$.

Appendix II: Step 2 in Proof of Proposition 3

Let G be the lever DAG $a \to x \to y$. Denote $p_{ay} \equiv p(x = 1 \mid a, y)$. Our objective is to find the maximal values for $p_G(y = 1 \mid a = 1)$ and $p_G(y = 1 \mid a = 0)$ subject to the constraint that either $p_{a^*1} = p_{a^*0} \in \{0, 1\}$ for some a^*, or $p_{1,y^*} = p_{0,y^*} \in \{0, 1\}$ for some y^*.

Recall that

$$p_G(y = 1|a = 1) = p(x = 1|a = 1)p(y = 1|x = 1)+p(x = 0|a = 1)p(y = 1|x = 0)$$

and by NSQD,

$$p_G(y = 1|a = 0) = \frac{\mu - \alpha p_G(y = 1|a = 1)}{1 - \alpha}$$

Since we are free to choose what outcome of x to label as 1 or 0, there are four cases to consider.

Case 1. Let $X_{a=1,x=1}$ be the set of lever variables that satisfy $p_{11} = p_{10} = 1$. It follows that for every $x \in X_{a=1,x=1}$, $p(x = 1|a = 1) = 1$ while $p(x = 0|a = 1) = 0$. Hence,

$$\max_{x \in X_{a=1,x=1}} p_G(y = 1|a = 1) = \max_{x \in X_{a=1,x=1}} p(y = 1|x = 1)$$

and

$$\max_{x \in X_{a=1,x=1}} p_G(y = 1|a = 0) = \frac{\mu - \alpha \min_{x \in X_{a=1,x=1}} p_G(y = 1|x = 1)}{1 - \alpha}$$

where

$$p(y = 1|x = 1) = \frac{\alpha\mu + (1 - \alpha)\mu p_{01}}{\alpha\mu + (1 - \alpha)\mu p_{01} + \alpha(1 - \mu) + (1 - \alpha)(1 - \mu)p_{00}}$$

The R.H.S. of this equation is maximized when $p_{01} = 1$ and $p_{00} = 0$, and it is minimized when $p_{01} = 0$ and $p_{00} = 1$. Therefore,

$$\max_{x \in X_{a=1,x=1}} p_G(y = 1|a = 1) = \frac{\mu}{\mu + \alpha(1 - \mu)}$$

where this maximum is attained by $p_{11} = p_{10} = p_{01} = 1$ and $p_{00} = 0$ (which is equivalent to a lever variable defined as $x = y + a(1-y)$, while

$$\max_{x \in X_{a=1,x=1}} p_G(y = 1|a = 0) = \frac{\mu - \alpha\frac{\alpha\mu}{\alpha+(1-\alpha)(1-\mu)}}{1-\alpha} = \frac{\mu(\alpha + 1 - \mu)}{1 - \mu(1-\alpha)}$$

where this maximum is attained by $p_{11} = p_{10} = p_{00} = 1$ and $p_{01} = 0$ (which is equivalent to a lever variable defined as $x = a + (1-a)(1-y)$).

Case 2. Let $X_{a=0,x=0}$ be the set of lever variables that satisfy $p_{01} = p_{00} = 0$. Hence,

$$\max_{x \in X_{a=0,x=0}} p_G(y = 1|a = 0) = \max_{x \in X_{a=0,x=0}} p(y = 1|x = 0)$$

and by NSQD,

$$\max_{x \in X_{a=0,x=0}} p_G(y = 1|a = 1) = \frac{\mu - (1-\alpha)\min_{x \in X_{a=0,x=0}} p(y = 1|x = 0)}{\alpha}$$

where

$$\begin{aligned} p(y &= 1|x = 0) = \frac{\alpha\mu(1-p_{11}) + (1-\alpha)\mu}{\alpha\mu(1-p_{11}) + (1-\alpha)\mu + \alpha(1-\mu)(1-p_{10}) + (1-\alpha)(1-\mu)} \\ &= \frac{1}{1 + \frac{\alpha(1-\mu)(1-p_{10})+(1-\alpha)(1-\mu)}{\alpha\mu(1-p_{11})+(1-\alpha)\mu}} \end{aligned}$$

Since the R.H.S. of this equation *decreases* in p_{11} and *increases* in p_{10} we have that

$$\max_{x \in X_{a=0,x=0}} p_G(y = 1|a = 0) = \frac{\mu}{\mu + (1-\alpha)(1-\mu)}$$

which is attained by $p_{01} = p_{00} = p_{11} = 0$ and $p_{10} = 1$ (which is equivalent to a lever variable $x = a(1-y)$), while

$$\max_{x \in X_{a=0,x=0}} p_G(y = 1|a = 1) = \frac{\mu - (1-\alpha)\frac{(1-\alpha)\mu}{(1-\alpha)\mu+(1-\mu)}}{\alpha} = \frac{\mu(2 - \alpha - \mu)}{1 - \alpha\mu}$$

which is attained by $p_{01} = p_{00} = p_{10} = 0$ and $p_{11} = 1$ (which is equivalent to a lever variable $x = ay$).

Case 3. Let $X_{y=1,x=1}$ be the set of lever variables that satisfy $p_{01} = p_{11} = 1$.

Hence,

$$\max_{x \in X_{y=1,x=1}} p_G(y=1|a=1) = \max_{x \in X_{y=1,x=1}} p(x=1|a=1)p(y=1|x=1)$$

and by NSQD,

$$\max_{x \in X_{a=0,x=0}} p_G(y=1|a=0) = \frac{\mu - \alpha \min_{x \in X_{y=1,x=1}} p(x=1|a=1)p(y=1|x=1)}{1-\alpha}$$

where for $x \in X_{y=1,x=1}$,

$$p(x=1|a=1)p(y=1|x=1) = (\mu+(1-\mu)p_{10}) \cdot \frac{\mu}{\mu+\alpha(1-\mu)p_{10}+(1-\alpha)(1-\mu)p_{00}}$$

Since the R.H.S. of this equation is *increasing* in p_{10} and *decreasing* in p_{00} it follows that

$$\max_{x \in X_{y=1,x=1}} p_G(y=1|a=1) = \frac{\mu}{\mu+\alpha(1-\mu)}$$

which is attained by $p_{01} = p_{11} = p_{10} = 1$ and $p_{00} = 0$ (which is equivalent to a lever variable $x = y + a(1-y)$) whereas,

$$\min_{x \in X_{y=1,x=1}} p_G(y=1|a=1) = \frac{\mu^2}{\mu+(1-\alpha)(1-\mu)}$$

which is attained by $p_{01} = p_{11} = p_{00} = 1$ and $p_{10} = 0$ (which is equivalent to a lever variable $x = y + (1-y)(1-a)$) such that

$$\max_{x \in X_{a=0,x=0}} p_G(y=1|a=0) = \frac{\mu}{\mu+(1-\alpha)(1-\mu)}$$

Case 4. Let $X_{y=0,x=0}$ be the set of lever variables that satisfy $p_{00} = p_{10} = 0$. Maximizing $p_G(y=1|a=1)$ is equivalent to minimizing $1-p_G(y=0|a=1)$. Since $p(y=0|x=1) = 0$ it follows that

$$p_G(y=0|a=1) = p(x=0|a=1)p(y=0|x=0)$$

where

$$p(x = 0|a = 1) = \mu(1 - p_{11}) + (1 - \mu) = 1 - \mu p_{11}$$

and

$$p(y = 0|x = 0) = \frac{1 - \mu}{1 - \mu + \alpha\mu(1 - p_{11}) + (1 - \alpha)\mu(1 - p_{01})} = \frac{1 - \mu}{1 - \mu(\alpha p_{11} + (1 - \alpha)p_{01})}$$

Hence, we want to find p_{11} and p_{01} that minimize

$$\frac{(1 - \mu)(1 - \mu p_{11})}{1 - \mu(\alpha p_{11} + (1 - \alpha)p_{01})}$$

This expression *increases* in p_{01} and *decreases* in p_{11}. Therefore,

$$\max_{x \in X_{y=0,x=0}} p_G(y = 1|a = 1) = 1 - \frac{(1 - \mu)^2}{1 - \alpha\mu} = \frac{\mu(2 - \alpha - \mu)}{1 - \alpha\mu}$$

which is attained by $p_{10} = p_{00} = p_{01} = 0$ and $p_{11} = 1$ (which in turn is equivalent to a lever variable $x = ay$)

Similarly,

$$\max_{x \in X_{y=0,x=0}} p_G(y = 1|a = 0) = 1 - \min_{x \in X_{y=0,x=0}} p_G(y = 0|a = 0)$$

where $p_G(y = 0|a = 0)$ is equal to

$$p(x = 0|a = 0)p(y = 0|x = 0) = \frac{(1 - \mu)[(1 - \mu) + \mu(1 - p_{01})]}{(1 - \mu) + (1 - \alpha)\mu(1 - p_{01}) + \alpha\mu(1 - p_{11})}$$

Since the R.H.S. of this expression *decreases* in p_{01} and *increases* in p_{11}, we have that

$$\max_{x \in X_{y=0,x=0}} p_G(y = 1|a = 0) = 1 - \frac{(1 - \mu)^2}{1 - \mu(1 - \alpha)} = \frac{\mu(1 + \alpha - \mu)}{1 - \mu(1 - \alpha)}$$

which is attained by $p_{10} = p_{00} = p_{11} = 0$ and $p_{01} = 1$ (which is equivalent to a lever narrative $x = y(1 - a)$).

From the above four cases we obtain two candidate lever variables for maximizing $p_G(y = 1|a = 1)$: $x = ay$ and $x = y + a(1 - y)$. The latter leads to a higher expected anticipatory payoff if and only if

$$\frac{\mu}{\mu + \alpha(1-\mu)} > \frac{\mu(2-\alpha-\mu)}{1-\alpha\mu}$$

which holds if and only if $\mu < 1 - \alpha$. Similarly, we obtain two candidate lever variables for maximizing $p_G(y = 1|a = 0) : x = y(1-a)$ and $x = y + (1-y)(1-a)$. The latter leads to a higher expected anticipatory payoff if and only if

$$\frac{\mu}{\mu + (1-\alpha)(1-\mu)} > \frac{\mu(1+\alpha-\mu)}{1-\mu(1-\alpha)}$$

which holds if and only if $\mu < \alpha$.