The Political Economy of Morality Policies: Collective Choice and Denial*

Vincent Anesi† Renee Bowen‡

March 20, 2025

Abstract

Preferences over some policies require moral judgements that are influenced by an individual's deeply held worldviews. Political scientists have sometimes referred to such policies as *morality policies*. Examples include climate change, humanitarian crises, or discrimination. While preferences for morality policies are shaped by worldviews, these worldviews can themselves respond to the policy environment, many times resulting in denial. This paper develops a theory of how collective choice institutions influence the formation of preferences over addressing such problems. We present a model in which agents can adopt worldviews to mitigate the psychological discomfort of acknowledging a problem, weighing the benefits of future policy actions against the costs of recognizing unpleasant realities. We identify necessary and sufficient conditions for the policy action to be politically sustainable, and show that anonymous voting rules, such as majority voting, are conducive to denialist free riding where agents delegate the burden of acknowledging the problem to others. The model also predicts that collective decision-making generally amplifies denialist behavior and generates polarization of worldviews when action is implemented. These findings shed light on how the design of collective choice institutions affects both policy outcomes and individuals' incentives to face uncomfortable truths.

*We are grateful to Raphaël Levy for helpful comments and suggestions.

†,University of Luxembourg, Email: vincent.anesi@uni.lu

‡University of California, San Diego, Email: trbowen@ucsd.edu

1 Introduction

> *"If we deny reality, if we don't feel the pain of what is happening in Bosnia or in the house next door, we don't act, even in the small ways available to us, to change those realities. Instead, we may construct alternatives to reality, fantasy-worlds that further betray our perception of what is really occurring. Nations, in fact, operate on the basis of such shared reconstructions of reality much of the time. This collective fantasy life is both revealed in politics and supported and maintained by it.*[1] "

The impact of agents' preferences on collective choices is well-studied, but much less is known about how collective choice institutions may affect the formation of these preferences. In this paper, we take up this question. Specifically, we study the formation of preferences over *morality policies*—those influenced by not only instrumental policy impact, but also by deeply held worldviews. Mooney (1999) argues that a "morality policy and its politics are characterized by debates over first principles, in which 'at least one advocacy coalition involved has portrayed the issue as one of morality [...] and used moral arguments in its policy advocacy' ". Such moral arguments often result in a clash of values that cannot be resolved by argument. To understand the collective choice outcome for a morality policy, an agent's worldview must be accounted for in their preferences. In this paper we theoretically explore how collective action on morality policies can shape individuals' worldviews, and in turn impact the possibility of collective action to address such policies.

One category of morality policies are those that do not affect the individual's present conditions, but that of future generations. Climate change is a striking example. World leaders and scientists formally recognized it as a critical environmental issue as early as 1979 during the First World Climate Conference. However, its existence faced fierce contestation for over forty yearrs (Hale, 2024, Norgaard, 2011). Markowitz and Shariff (2012) identify psychological factors that affect an individual's moral judgement of climate change, including: that uncertainty about its impact breeds wishful thinking; guilt bias; abstractness and complexity. Other examples in this category include public debt, the protection

[1] M.A. Milburn and S. Conrad in *The Politics of Denial,* 1996.

of rain forests, and the insolvency of social-security systems.

Another category of morality policies are those that do not affect the individual personally but rather other individuals with whom they do not have a particular connection. Examples include humanitarian crises or use of force in foreign countries (e.g., conflicts, famines, earthquake devastation). Finnemore (2003) documents how collective beliefs and values about such interventions change over time, and what was once heterodox, becomes completely "obvious" and is widely supported. This suggests that values related to morality policies can change over time. Acknowledgement of the need for action on morality policies often induces psychological discomfort because existing worldviews are challenged. Psychologist divide stress coping strategies into two categories: problem-focused coping—efforts to change or manage the stressor; and emotion-focused coping—managing emotional distress rather than altering the situation (Lazarus and Folkman, 1984). That is, changing worldviews to make the political context more "comfortable" is an important feature of morality policies.

It is a well established fact in psychology that individuals employ various self-deception mechanisms (such as denial, rationalization, repression, etc.) to evade certain unpleasant elements of reality, even when faced with overwhelming evidence to the contrary. In his pioneering work on consciousness and perception, James (1890) famously discussed the concepts of "selective attention" and "avoidance." He examined how individuals may selectively avoid certain aspects of reality, particularly those that may be distressing or conflicting with their desires, and the role of attention in shaping their perception of the world around them. In fact, Nietzsche, as early as 1886, also discussed the notion of self-deception in his philosophical writings, discussing the human inclination towards "willful ignorance" and the avoidance of uncomfortable truths or realities. As Milburn and Conrad (1996) more recently put it in a political-economic context, "we are all capable of failing to see or hear the pain and evil around us or—when it becomes glaringly apparent and the evening news shows us corpses tumbling from the back of a dump truck into a mass grave in Rwanda—of failing to appreciate the emotional significance of what we see, of connecting the image with reality [...]." .

A critical aspect in addressing morality issues, thus, lies in whether agents' *choose* worldviews that lead to the acknowledgement of a problem or not. We consider that agents can endogenously adopt worldviews that shape how they evaluate future policy choices. Agents, therefore, have an opportunity to adjust their world view and choose to ignore a problem if it is too uncomfortable. A consequence of denial is that the incentive to take necessary actions to address the problem vanishes. For example, denial of climate change has long been considered as one of the main obstacles to transitioning to a green economy (e.g., Norgaard, 2011; Mann and Toles, 2016; or Ekberg et al., 2022). The underlying logic is straightforward: if a decisive group of agents espouse denialist worldviews, then they oppose corrective reforms, and political institutions in turn aggregate those policy preferences into an agenda of inaction. The goal of this paper is to uncover subtler channels through which the causality also flows from the political process to the formation of worldviews: the institution (be it an electorate, a committee, a legislature, etc.) will affect the incentives of pivotal members to deal with uncomfortable truths.

Our analysis rests on four important premises:

(P1) *Action to address moral policy problems are costly.* As one would expect, policy reforms may imply economic costs, such as increased tax burdens. But costs need not only be economic. Action could also involve a toll on well-being, such as changes in one's lifestyle, or even loss of life, as in the case of interventions in foreign conflicts.

(P2) *If an individual acknowledges the existence of a problem, then taking action to address the problem will yield her benefits.* One way to think about this is increasing the individual's expected future utility by addressing a known problem. It can also account for someone who recognizes the dangers posed by a warming climate feeling reassured about her own and her children's future when environmental reforms are being implemented. Such benefits may not be related to immediate economic rewards. Individuals often experience a "warm glow," an inherent satisfaction from supporting prosocial policies that benefit the common good, feeling a sense of moral rectitude in "doing the right thing" (e.g., Andreoni, 1989; Bénabou and Tirole, 2006; or Dunn et al., 2008).

(P3) *Recognizing the existence of a moral problem causes psychological discomfort.* This is one of the mainstays of the aforementioned psychology literature on denial; it is the factor that drives individuals toward self-deception. For example, there is ample evidence that the degradation of the environment, including issues such as climate change, pollution, and resource depletion, can generate significant anxiety and distress, as well as feelings of responsibility, about the future of the planet and the well-being of future generations (e.g., Kals and Maes, 2002; Whitmarsh and Capstick, 2018; or Clayton and Karazsia, 2020); concerns about the stability and sustainability of pension systems can lead to feelings of financial insecurity and anxiety about retirement planning (e.g., Olsen and Aas, 2019); and images and reports of suffering and devastation in foreign conflicts make individuals experience discomfort and distress, grappling with the ethical implications of those conflicts and the suffering of those affected. By ignoring the existence of the problem, the individual is using emotion-focused coping to relieve the cost of such discomfort.

(P4) *When forming future worldviews, agents give weight to their current and future perspectives.* This is the main building block of Bernheim et al.'s (2021) theory of endogenous worldview formation. In settings with changing worldviews, agents can evaluate future policy choices through two distinct lenses: one based on their present worldview, and the other based on the worldview they expect to hold when those policies are chosen and implemented. A personality trait, which they call "mindset flexibility," determines the relative weights an individual attaches to their current and future outlooks (or "selves").

We present a two-stage model in which a particular social problem has arisen. Agents initially inherit different worldviews about its existence—these initial worldviews may stem from personal experiences, parental upbringing, social influences, etc. Agents have an opportunity to adjust their worldview in the first stage, prior to making a collective decision on taking an action to address the problem in the second stage. When choosing worldviews agents take into consideration the consequences that their newly formed worldviews will have on their voting behavior, both from their current and future perspectives—our premise (P4)—and the psychological discomfort associated with acknowledging the problem—premise (P3).

There are two main types of agents: *pragmatists,* for whom the costs (P1) and benefits (P2) of the policy action dominate the psychological discomfort (P3) caused by the problem; and *idealists,* for whom the psychological discomfort instead dominates. Conditional on the action being implemented, the former are better off convincing themselves that the problem exists, and that the associated costs are justified; the latter are always better off convincing themselves that the problem is a non-issue, thus no longer having to worry about it.

Since the psychological discomfort associated with acknowledging the problem is so burdensome for an idealist, what prevents her from systematically adopting a worldview of complete denial of the problem, thereby freeing herself from all the anxieties and concerns it may cause? The answer lies in the fact that individuals are not completely mindset flexible: the idea that her future self, in complete denial of the problem, will do nothing to address it concerns the current self of an idealist who may initially acknowledge its existence. As demonstrated by Bernheim et al. (2021) in *individual* decision-making settings, mindset inflexibility thus confronts an idealist with a dilemma when adjusting her worldview: satisfying her future self by denying the existence of the problem, or satisfying her current self by sufficiently acknowledging it to remain willing to confront it. Thus, while pragmatists tend to lean towards extreme worldviews—fully acknowledging the problem if they anticipate enforcing the action, and completely denying it if they anticipate maintaining the inaction status quo—their mindset inflexibility drives the idealists towards moderation in their worldview choices.

The analysis of our model reveals that in situations of *collective* decision making, such as those of interest here, some voting rules encourage idealists to try to escape that dilemma, with significant consequences for the political sustainability of policy action on morality policies and individual worldview formation. For example, in the classic case where the collective choice to adopt or not the policy action is decided by a majority vote, an idealist can now hope to rely on other agents to continue worrying about the problem and form a majority coalition that will enforce the action, while at the same time, she may herself engage in denial. Her current self, aware of the problem's existence, is thus assured of

the implementation of the action, while her future self will be freed from the psychological discomfort. This can be thought of as a form of "denialist free riding" on other agents' acknowledgement of the problem that benefits all the pro-action idealists' initial selves, whether they choose to continue to recognize its existence or not.

We establish two necessary and sufficient conditions for policy action to be sustainable in equilibrium—*decisiveness* and *no free riding*. These two conditions shed light on a tension regarding the stringency of the voting rule: on the one hand, decisiveness requires that the rule be sufficiently permissive to make a pro-action coalition powerful enough to enforce the action; on the other hand, no free riding requires that the rule be sufficiently stringent to curb denialist free-riding opportunities. What are the voting rules that can resolve this tension? Our equilibrium analysis reveals that a critical factor in the enforceability of the action is lack of *anonymity* of the voting rule. Anonymity, which occupies a fundamental place in social choice theory, is often deemed desirable because it requires all agents to be treated equally in preference aggregation. Anonymous voting rules, i.e., quota rules, are thus frequently encountered in real-world collective decision making. However, our equilibrium analysis shows that this property also makes all agents perfectly substitutable in forming the decisive coalition meant to enforce policy action, which in turn fosters denialist free riding by idealists who hope to rely on other agents to address the problem on their behalf. In fact, we prove that under mild conditions on the distribution of agents' preferences, it is simply impossible to sustain policy action in equilibrium under a quota rule, *regardless of the value of the quota,* and even if the action is highly desirable from a social welfare perspective.

For a rule maker aiming to promote the implementation of the action, the solution necessarily involves granting veto power to idealists whose free riding prevents the action from being enforceable under anonymous rules. Free riding is then no longer an option for these idealist vetoers, whose vote is now crucial for the passage of the action; they are no longer able to delegate the responsibility of addressing the problem to others. They have therefore no other choice but to ensure that their future selves remain sufficiently aware of the problem to vote in favor of the action at the voting stage. This illustrates how political

power can have adverse consequences for these idealists. This stands in sharp contrast to standard political-economic settings with exogenous worldviews, where political power is always beneficial.

Having characterized equilibrium policy outcomes in the voting stage, we then turn to individual worldview choices in the first stage. An asymmetry arises here between equilibria in which the policy action is enforceable and those where it is not: in the former, all pragmatists choose to fully acknowledge the existence of the problem, while idealists seek to deny it as much as possible, resulting in polarization of worldviews; in the latter, all agents unanimously adopt a worldview of complete denial. Thus, a polarization of worldviews arises endogenously when the action is implemented in equilibrium, and conversely, a complete alignment of worldviews occurs when inaction is maintained. In fact, our model shows that the collective decision-making process generally leads individuals toward more denialist behavior than if they were making decisions individually.

Related Work

Before proceeding further, it is important to highlight the distinctions between this paper and the closely related political-economy literature on policy preference formation. The existing political-economy literature on self-deception has largely focused on the awareness-management framework introduced by Bénabou and Tirole (2002). This framework has been widely applied to various policy-making contexts, such as redistributive politics (Bénabou and Tirole, 2006), the scope of government (Bénabou, 2008), and political agency (Levy, 2014).[2] A common theme in this literature is the strategic complementarities that emerge in equilibrium between policy choices and individual self-deception. For instance, in the model by Bénabou and Tirole (2006), voters who expect little income redistribution choose to "forget" pessimistic signals about long-term returns to effort in order to maintain their motivation to work hard for high future incomes, essentially investing in a

[2]We refer the reader to Bénabou and Tirole (2016) for a more detailed account of that literature and the broader literature on self-deception in organizations. Golman et al (2017) also survey the literature on information avoidance.

"just-world ideology." This ideology, in turn, leads them to support policies of low income redistribution. Such self-deception becomes even more valuable when perceptions of effort costs diverge between their present and future selves. This complementarity between denial choices and equilibrium policies is also found in the behavior of pragmatists in our model or morality policies. However, a key contribution of our paper is to shed light on a different category of agents, the idealists, whose self-deception and voting incentives exhibit instead a form of substitutability not captured by existing models. The anticipation of a policy reform to address a problem reinforces an idealist's equilibrium incentives to deny the existence of that problem. What distinguishes idealists from pragmatists is the magnitude of psychological discomfort induced by acknowledging the moral problem.

Idealists play an important role in our analysis. Oour goal is to better understand the interaction between various institutional rules that govern decision-making and worldview formation when denial features prominently in agents' voting incentives. These rules matter for two main reasons: first, because they determine under what conditions agents will be endogenously pivotal when choosing policy; and second, because they can generate polarization depending on the initial distribution of worldviews and payoffs. Neither the heterogeneity of interests among agents nor pivotality plays a role in the above literature since these models typically assume there is a continuum of agents who are ex ante identical, hence never pivotal.[3]

In our context, worldview formation refers to degrees of denial regarding unpleasant, painful, or frightening aspects of reality, which is a deliberate (yet often unconscious) *choice* (e.g., Norgaard, 2011; Cohen, 2013). This naturally leads us to adopt the model of chosen preferences developed by Bernheim et al.'s (2021) — who concentrate on *individual* decision-making. This differs from other political economic models that consider worldviews as beliefs about the underlying processes shaping policy outcomes, such as Levy et al. (2022) and Izzo et al. (2023). In those models, a worldview constitutes a "theory of

[3]There is ex ante heterogeneity between two homogeneous groups, the poor and the rich, in Bénabou and Tirole (2006). However, the only role of this heterogeneity is to make the redistributive policy operational. The key determinant of equilibrium outcomes is whether self-signaling incentives will drive enough poor voters to support either a high or low income tax.

how the world works" that influences voting behavior in political competition settings.

A vast literature studies policy inertia and has attributed it to a variety of factors. These include, uncertainty regarding the identity of reform beneficiaries (e.g., Fernandez and Rodrik, 1991; Strulovici, 2010), asymmetric costs disproportionately borne by those initiating policy changes (e.g., Alesina and Drazen, 1991), mutual fear among policy parties of unfavorable concessions reinforcing status-quo deadlock (e.g., Howitt and Wintrobe, 1995), policy-induced private actions entrenching future political support for the status quo (e.g., Coate and Morris, 1999), endogenous status-quo dynamics in political bargaining (e.g., Anesi and Seidmann, 2015; Dziuda and Loeper, 2016), or the interplay between veto rights and redistributive constraints in policy experimentation (e.g., Anesi and Bowen, 2021). Such policies preferences are primarily driven by instrumental incentives. Our focus on morality policies and collective denial introduces a behavioral dimension to the study of policy intertia.

Our model shows how the interplay between collective decision-making and individual worldview formation can lead to extreme polarization among voters. In this respect, the paper relates to the recent and growing political-economy literature that examines how voters' endogenous information processing may induce polarization of policy preferences. However, in those models, political polarization arises through different channels, such as the allocation of attention (Yuksel, 2022; Nunnari and Zápal, forthcoming), the selection of information sources (Perego and Yuksel, 2022), costly memory (Eguia and Hu, 2023) or limited temporal memory (Levy and Razin, 2025), and information sharing with personal connections (Bowen et al., 2023), to name a few.

2 Model

Basics. Consider a group of agents (a committee, electorate, legislature, etc.), $N \equiv \{1, \ldots, n\}$, confronted with a morality policy problem whose consequences are not immediate (e.g., civil conflict in a foreign country, climate change, etc.). At the beginning of the game, members of the group are endowed with heterogeneous levels of recognition of the problem: each individual i initially holds a worldview $\alpha_i^0 \in [0, 1]$ that it exists (or

"matters"), which we refer to as her *inherited worldview.* These inherited worldviews are assumed to have been acquired from past personal experiences, parental upbringing, social influences, and similar factors. The game then comprises two stages: a worldview-choice stage, and a voting stage. During the worldview choice stage, each agent has an opportunity to change her inherited world-view to some new worldview $\alpha_i \in [0,1]$.[4] Holding new worldviews $\alpha = (\alpha_1, \ldots, \alpha_n)$, in the voting stage agents simultaneously vote in favor of an alternative from $\{a, b\}$, where $x = a$ means that they chose to act on the problem, while $x = b$ means that they choose not to act.

Preferences. Adopting the Bernheim et al. (2021) framework, we assume that the voting stage-payoffs of agent i with worldview α_i from choosing alternative x are represented by the convex combination

$$U_i(\alpha_i, x) \equiv \alpha_i u_i(x) + (1 - \alpha_i) v_i(x) \ . \tag{1}$$

The function $u_i \colon \{a, b\} \to \mathbb{R}$ can be thought of as the policy preferences of each agent i when she fully recognizes the existence of the problem (i.e., $\alpha_i = 1$), while $v_i \colon \{a, b\} \to \mathbb{R}$ are her preferences when she fully denies it (i.e., $\alpha_i = 0$). We make the following natural assumptions on u_i and v_i.

Assumption 1.

1. *Each agent i is better off acting on an acknowledged problem, i.e., $u_i(b) < u_i(a)$.*
2. *Each agent i is better off not acting when there is no problem, i.e., $v_i(a) < v_i(b)$.*
3. *Each agent i is better off when there is no problem and no action is taken than when there is a problem and the action is taken, i.e., $u_i(a) < v_i(b)$.*

[4] To streamline the presentation of our results, we do not explicitly model the process through which agents self-deceive, allowing them to choose any worldview in $[0, 1]$ without constraint. One alternative is a belief updating framework that models self-deception as in Bénabou's (2013). We show in Section C of the Appendix, however, that our results would remain unchanged if we use the Bénabou's (2013) framework, despite the constraints it imposes on the agents' latitude to self-deceive.

As discussed in the Introduction, in a model with endogenous worldviews, it is important to distinguish between the preferences of agents before and after they adopt their new worldviews (recall P4). At the voting stage, an agent i has endogenous worldview α_i and, consequently, seeks to enforce the alternative x that maximizes $U(\alpha_i, x)$. When it comes to choosing her worldview at the initial stage of the model, the agent takes into account both the payoff of her future "self" at the time of voting, and the payoff given her inherited worldview α_i^0. Her payoff is then given by

$$\lambda_i U_i(\alpha_i^0, x) + (1 - \lambda_i) U_i(\alpha_i, x) \ , \tag{2}$$

where $\lambda_i \in (0, 1)$ is a parameter that reflects her "mindset (in)flexibility" (see Bernheim et al., 2021). It captures the agent's balancing act between her present and future selves' interests: while she cares about her future well-being, she also takes into account the policies that will be implemented in the future according to her current worldview. A higher value for λ_i corresponds with less mindset flexibility because more weight is placed on the agent's current world view.

Decision rules. Collective decision making is governed by a *voting rule,* formally modeled as a collection $\mathcal{D} \subseteq 2^N \setminus \{\emptyset\}$ of *decisive coalitions.* The status quo b is changed to a if and only if a coalition in $\mathcal{D}$ votes in favor of the latter alternative. We make standard assumptions on $\mathcal{D}$ (e.g., Austen-Smith and Banks, 1999). We assume the voting rule $\mathcal{D}$ is *proper,* i.e., every pair of decisive coalitions has a nonempty intersection: $C_1, C_2 \in \mathcal{D}$ implies $C_1 \cap C_2 \neq \emptyset$. We further assume $\mathcal{D}$ is *monotonic,* i.e., any superset of a decisive coalition is itself decisive: $C_1 \in \mathcal{D}$ and $C_1 \subseteq C_2$ imply $C_2 \in \mathcal{D}$.

Nontrivial equilibrium. We study (pure-strategy) subgame perfect equilibria (SPEs) of the game where: in the voting stage, for any worldview profile α inherited from the previous stage, each agent i seeks to maximize (1); and in the worldview-choice stage, each agent i seeks to maximize (2). As is common in the voting literature, we rule out implausible behavior in voting stages by restricting attention to SPEs that induce undominated strategies in those stages. Moreover, in order to limit the number of possible cases (without

affecting the paper's conclusions), we will further assume that in case of a tie, an agent will always favor alternative a over b. In what follows, any reference to an *equilibrium* is to an undominated SPE that satisfies this tie-breaking condition.

Coordination failures analogous to those in voting stages also arise in worldview-choice stages, but cannot be ruled out by eliminating weakly dominated strategies stage-wise. We provide an example. We will say that $\mathcal{D}$ is *collegial* if there is some agent who has a veto, i.e., $\bigcap \mathcal{D} \neq \emptyset$. The family of collegial voting rules includes unanimity rule (i.e., $\mathcal{D} = \{N\}$) and dictatorships (i.e., $\mathcal{D} = \{C \subseteq N \colon C \ni i\}$ for some $i \in N$. In contrast, any quota rule defined by $\mathcal{D} = \{C \subseteq N : |C| \geq q\}$, where q satisfies $n/2 < q < n$, is non-collegial. Whenever $\mathcal{D}$ is noncollegial, if an agent expects all the others to adopt worldview 0 in every period, then her only best response is to adopt 0 too, regardless of her current worldview. Similarly, if the pragmatists constitute a (non-minimal) decisive coalition, then there is always an equilibrium in which they all adopt worldview 1, even if they all have an initial worldview of 0. We seek to eliminate such trivial equilibria.

An equilibrium will be referred to as "nontrivial" if it is induces strong Nash equilibria stage-wise—this is in the spirit of what DeMarzo (1992) calls a "sustainable social norm." More specifically, let σ be a strategy profile; and, for all $\alpha \in [0,1]^n$, let $\chi^\sigma(\alpha) \in \{a, b\}$ be the alternative chosen (under σ) in the voting stage if the agents hold worldviews $\alpha = (\alpha_1, \ldots, \alpha_n)$. Then, we say that an equilibrium σ is *nontrivial* if the worldview profile it prescribes in the first stage is a strong Nash equilibrium of the normal-form game described by the payoff functions $\pi_i^\sigma \colon [0,1]^n \to \mathbb{R}$, where

$$\pi_i^\sigma(\alpha) \equiv (1 - \lambda_i) U_i\big(\alpha_i, \chi^\sigma(\alpha)\big) + \lambda_i U_i\big(\alpha_i^0, \chi^\sigma(\alpha)\big) \ ,$$

for all $i \in N$ and $\alpha \in [0,1]^n$.

3 Pragmatists and Idealists

Two essential first steps in characterizing the equilibria of the model presented above are: (i) analyzing how worldview choices in the first stage affect voting behavior in the second, and (ii) understanding the individual incentives that drive equilibrium worldview choices.

The latter step is the focus of Bernheim et al. (2021), who emphasize the distinction between two types of agents, which we will refer to as *pragmatists* and *idealists*. The collective decisions we study will hinge on the interaction between these two agents.

Voting. The equilibrium voting behavior of any agent i in the second stage of the game is straightforward: having adopted a worldview α_i in the first stage, the agent votes to act on the problem if and only if she is sufficiently convinced of its reality. That is, she votes in favor of action a if and only if $U_i(\alpha_i, b) \leq U_i(\alpha_i, a)$, or equivalently,

$$\alpha_i \geq \frac{v_i(b) - v_i(a)}{v_i(b) - v_i(a) + u_i(a) - u_i(b)} \equiv \bar{\alpha}_i \ .$$

It follows that after agents have chosen worldviews $\alpha = (\alpha_1, \ldots, \alpha_n)$, then the equilibrium policy choice $\chi^\sigma(\alpha)$ is the same in all equilibria, i.e.,

$$\chi^\sigma(\alpha) = \begin{cases} a & \text{if } \{i \in N \colon \alpha_i \geq \bar{\alpha}_i\} \in \mathcal{D} \ , \\ b & \text{otherwise,} \end{cases}$$

for all $\alpha \in [0,1]^n$.

Worldview choice incentives. Assumption 1 implies

$$u_i(b) < u_i(a) < v_i(b) \ , \tag{3}$$

for each $i \in N$, indicating that $v_i(b)$ is the highest payoff for all agents. However it places no restriction on the ordering of $u_i(a)$ and $v_i(a)$. We say that agent i is a *pragmatist* if $u_i(a) > v_i(a)$ because it is always in her interest to have a worldview aligned with the chosen policy alternative. If $u_i(a) < v_i(a)$, we say that i an *idealist.* Regardless of the policy taken, idealists prefer to deny that the problem exists, as acknowledging it would cause them significant psychological discomfort.

To see why pragmatists and idealists matter, first consider agent i's worldview choice α_i if she is a dictator, so that the voting outcome $\chi^\sigma(\alpha_i, \alpha_{-i})$ only depends on α_i. By backward induction, she seeks to maximize

$$\lambda_i U_i\big(\alpha_i^0, \chi^\sigma(\alpha_i, \alpha_{-i})\big) + (1 - \lambda_i) U_i\big(\alpha_i, \chi^\sigma(\alpha_i, \alpha_{-i})\big) \ .$$

Substituting (1) and rearranging terms, this expression can be rewritten as

$$v_i\big(\chi^\sigma(\alpha_i,\alpha_{-i})\big) + \big[\lambda_i\alpha_i^0 + (1-\lambda_i)\alpha_i\big]\Big[u_i\big(\chi^\sigma(\alpha_i,\alpha_{-i})\big) - v_i\big(\chi^\sigma(\alpha_i,\alpha_{-i})\big)\Big] \ ,$$

which better highlights the relationship between λ_i, α_i^0, and the difference between $u_i(a)$ and $v_i(a)$. The term $\lambda_i\alpha_i^0 + (1-\lambda_i)\alpha_i$ captures the trade-off between agent i's inherited and chosen worldviews. This trade-off is an important driver of worldview choice when the agent is very mindset inflexible, i.e, λ_i is high.

Suppose λ_i is small, so that agent i is highly flexible in her mindset and places little importance on her current worldview α_i^0. Then, her payoff is maximized by setting α_i to zero and leading to the choice of $\chi^\sigma(0,\alpha_{-i}) = b$. This maximizes the first term and yields her highest payoff, $v_i(b)$. In this case, the ordering of $u_i(a)$ and $v_i(a)$ is inconsequential. Next, consider the case where λ_i is large, making agents mindset inflexible and thus strongly attached to their inherited worldview. If α_i^0 is close to zero, there is little conflict between agent i's present and future selves, so she will still choose $\alpha_i = 0$ and induce policy choice $\chi^\sigma(0,\alpha_{-i}) = b$ to maximize her payoff. However, if α_i^0 is large and the agent's current worldview acknowledges the problem, the difference between $u_i(a)$ and $v_i(a)$ becomes crucial.

First, suppose agent i is a *pragmatist* so that $u_i(a) > v_i(a)$. A large value of $\lambda_i\alpha_i^0$ implies that she puts a lot of weight on $u_i\big(\chi^\sigma(\alpha_i,\alpha_{-i})\big) - v_i\big(\chi^\sigma(\alpha_i,\alpha_{-i})\big)$ being positive. This requires that $\chi^\sigma(\alpha_i,\alpha_{-i}) = a$, which, in turn, implies that she should choose $\alpha_i = 1$ to maximize her payoff. Now suppose that $u_i(a) < v_i(a)$, so that agent i is an *idealist.* Assuming again that $\lambda_i\alpha_i^0$ is large (i.e., close to 1), i's payoff is close to $u_i\big(\chi^\sigma(\alpha_i,\alpha_{-i})\big)$, leading her to seek $\chi^\sigma(\alpha_i,\alpha_{-i}) = a$, similar to the pragmatist. As we have seen above, she can only induce this choice by selecting a worldview α_i greater than $\bar{\alpha}_i$. Unlike the pragmatists, however, $[u_i\big(\chi^\sigma(\alpha_i,\alpha_{-i})\big) - v_i\big(\chi^\sigma(\alpha_i,\alpha_{-i})\big)] = \big[u_i(a) - v_i(a)\big]$ is now negative. For a mindset inflexible idealist i whose worldview is initially high, it is therefore optimal to reduce her worldview only up to the cutoff value $\bar{\alpha}_i$ at which her future self will still prefer to implement action a. We formalize this intuition with the following proposition.

Proposition 0 (Bernheim et al., 2021)**.** *Suppose agent i is a dictator, then:*

(i) If i is a pragmatist, i.e., $i \in P \equiv \{i \in N : v_i(a) \leq u_i(a)\}$, she adopts worldview

$$\alpha_i^d \equiv \begin{cases} 0 & \text{if } \alpha_i^0 < \alpha_i^* \ , \\ 1 & \text{otherwise,} \end{cases}$$

where

$$\alpha_i^* \equiv \bar{\alpha}_i + \frac{1-\lambda_i}{\lambda_i} \frac{v_i(b) - u_i(a)}{u_i(a) - u_i(b) + v_i(b) - v_i(a)} \ ;$$

(ii) if i is an idealist, i.e., $i \in I \equiv \{i \in N : u_i(a) < v_i(a)\}$, she adopts worldview

$$\alpha_i^d \equiv \begin{cases} 0 & \text{if } \alpha_i^0 < \alpha_i^{**} \ , \\ \bar{\alpha}_i & \text{otherwise,} \end{cases}$$

where

$$\alpha_i^{**} \equiv \bar{\alpha}_i + \frac{1-\lambda_i}{\lambda_i} \frac{v_i(b) - U_i(\bar{\alpha}_i, a)}{u_i(a) - u_i(b) + v_i(b) - v_i(a)} \ .$$

The cutoffs α_i^* and α_i^{**} in Proposition 0 allow us to further distinguish between agents who are initially sufficiently convinced of the morality problem's existence and prefer that the action be taken, from those who are not sufficiently convinced, and favor inaction. We henceforth denote the *pro-action pragmatists* as P^a and the *pro-action idealists* as I^a. Formally,

$$P^a \equiv \{i \in P : \alpha_i^0 \geq \alpha_i^*\} \ \text{ and } \ I^a \equiv \{j \in I : \alpha_j^0 \geq \alpha_j^{**}\}.$$

In the next subsection, we build on the intuition developed here to characterize the conditions under which policy action can be sustained in equilibrium with more general voting rules. In particular, when agent i is no longer a dictator, her choice of worldview α_i must factor in the dependency of $\chi^\sigma(\alpha_i, \alpha_{-i})$ on the worldview choices α_{-i} of the other agents.

4 Equilibrium

We begin by considering the equilibrium behavior of pragmatists and idealists separately. The first result characterizes voting and worldview choices of pragmatists.

Lemma 1. *In any equilibrium, all the pragmatists $i \in P$ adopt the same worldview; and they all vote for the winning alternative. They adopt worldview* 1 *if the winning alternative is a, and worldview* 0 *otherwise.*

The proof of Lemma 1 is straightforward. First note that in the voting stage, each $i \in P$ always votes for the winning alternative. To see this, suppose without loss of generality that a is the equilibrium winning alternative and (towards a contradiction) that i votes for b. This implies that she adopted a worldview α_i smaller than $\bar{\alpha}_i < 1$ in the initial stage. But this in turn implies that she could profitably deviate by adopting instead worldview 1 in that stage, thus leading her future self to vote for a in the voting stage. By monotonicity of $\mathcal{D}$, alternative a would remain the winning alternative and since $U_i(\bar{\alpha}_i, a) < U_i(1, a)$, her future self would receive a higher payoff — a contradiction. Thus, she must vote for the winning alternative a in the voting stage and, consequently, adopt the worldview that maximizes $U_i(\cdot, a)$, i.e., $\alpha_i = 1$, in the first stage. By the same logic, each $i \in P$ first adopts worldview 0, and then votes for b, in any equilibrium where b is the winning alternative.

We now turn to the equilibrium behavior of the idealists.

Lemma 2. *In any equilibrium, each idealist $i \in I$ either adopts worldview $\bar{\alpha}_i$ and then votes for alternative a, or she adopts* 0 *and then votes for b.*

To see why Lemma 2 holds, first note that the voting behavior of any agent depends solely on her own worldview choice. It follows that if $i \in I$ initially adopts a worldview prompting her future self to vote for a, then her optimal choice must be $\alpha_i = \bar{\alpha}_i$; and conversely, if she adopts a worldview that induces her future self to vote for b, then her optimal choice must be $\alpha_i = 0$ — recall that $u_j(a) < v_j(a)$ and $u_j(b) < v_j(b)$, for all $j \in I$. Thus, in any equilibrium, idealists can be divided into two subsets: one in which they adopt worldview $\bar{\alpha}_i$ and vote for alternative a, and another where they adopt 0 and vote for b.

Lemmas 1 and 2 lead to our main result, fully characterizing equilibrium behavior.[5]

[5] A formal proof of Propostion 1 is provided in Section A of the Appendix.

Proposition 1. *For any voting rule $\mathcal{D}$, policy action a is the unique nontrivial equilibrium policy outcome if there exists a (possibly empty) subset $I^{a*} \subseteq I^a$ such that:*

(D) *the coalition $I^{a*} \cup P^a$ is decisive under $\mathcal{D}$; and*

(NFR) *for every $i \in I^{a*}$, the coalition $(I^{a*} \cup P) \setminus \{i\}$ is not decisive under $\mathcal{D}$.*

In such a case, idealists in I^{a*} adopts worldview $\alpha_i = \bar{\alpha}_i$; all other idealists adopt worldview 0; and all pragmatists adopt worldview 1. Furthermore, if there is no subset I^{a*} that satisfies conditions (D) *and* (NFR)*, then b is the unique equilibrium policy outcome, and all agents adopt worldview 0.*

The political sustainability of a policy action fundamentally hinges on the ability to form a coalition that satisfies both Conditions (D) and (NFR). The first of these conditions is familiar from traditional models of collective choice with exogenous preferences. It simply ensures the existence of a decisive coalition that supports the action. Condition (NFR), however, represents the main innovation of our model. As we will explain below, it ensures idealists are unable to "free ride" on the awareness of other agents.

The intuition for the interaction between Conditions (D) and (NFR) is as follows. In a nontrivial equilibrium, the worldviews α_i must constitute a strong Nash equilibrium of the worldview-choice stage. A coalition of pro-action idealists I^{a*} must satisfy Condition (D), otherwise, the pro-action pragmatists P^a could profitably deviate by all adopting worldview 0, inducing all the members of the decisive coalition $P \cup (I \setminus I^{a*})$ to vote for (and enforce) b. To see why (D) is not sufficient and (NFR) is required, first recall from Lemma 1 that, like the members of I^{a*}, all the pragmatists in P vote for a in an equilibrium where a is implemented. Suppose (NFR) is not satisfied, then there would be an i in I^{a*} such that $(I^{a*} \cup P) \setminus \{i\}$ is decisive. This agent i could profitably deviate by "free riding" on the other members of $I^{a*} \cup P$ as follows. By changing her equilibrium worldview choice in the first stage to 0, she does not affect her *current self's* utility, but would improve her *future self's* utility without affecting the outcome of the vote. This increases her first-stage payoff by

$$(1 - \lambda_i)\big[U_i(0, a) - U_i(\bar{\alpha}_i, a)\big] = (1 - \lambda_i)\bar{\alpha}_i\big[v_i(a) - u_i(a)\big] > 0.$$

Since this agent is an idealist, we know that the last expression is positive. Agent $i \in I^{a*}$ would thus have an incentive to deviate to worldview 0 in the first stage, leading to a contradiction.

This intuition demonstrates that the choice of a worldview shares characteristics with contributing to a public good. For any idealist who is initially convinced that the problem exists and must be addressed, the "contribution" is made by maintaining a worldview high enough to place her future self in the coalition that will enforce the policy action. This generates a utility benefit shared by *all* the agents who initially recognize the problem, even those who chose not to be part of the coalition that voted for a. Being part of the pro-action voting coalition bears an *individual* cost for an idealist, because it requires remaining aware of the problem's existence and enduring the psychological discomfort that accompanies it. She therefore has an incentive to let other agents (possibly other idealists) form this pro-action voting coalition and benefit from their voting choices without having to address the problem herself. In some cases, such "denialist free riding" can make the policy action politically unsustainable and lead the entire group to deny the existence of the problem.

The equilibrium characterization in Proposition 1 has several important implications, which we explore in Section 6. Before that, the next section provides a few illustrative examples to better understand how denialist free riding emerges, and how it interacts with collective-choice rules.

5 Illustrative Examples

We consider some examples with three agents. One of the three agents is an idealist, and the other two are pragmatists. That is, $I = \{1\}$ and $P = \{2, 3\}$. Suppose further that agents 1 and 2 initially recognize the existence of the problem, while agent 3 is in complete denial. That is, $\alpha_1^0 = \alpha_2^0 = 1$ and $\alpha_3^0 = 0$. Preferences are such that $\alpha_1^{**}, \alpha_2^*, \alpha_3^* \in (0, 1)$, so that 1 is the only pro-action idealist and 2 is the only pro-action pragmatist (i.e., $I^a = \{1\}$, $P^a = \{2\}$).

Pro-action idealist as dictator. Consider first the case where the voting rule $\mathcal{D}$ is dictatorial with agent 1 as the dictator; that is, $\mathcal{D} = \{C \subseteq N \colon C \ni 1\}$. Condition (D) is trivially satisfied with $I^{a*} = \{1\}$, so there is a decisive coalition to take the action in equilibrium. Moreover, given her role as a dictator, agent 1 can only rely on herself to enforce alternative a at the voting stage, thus Condition (NFR) is also satisfied. It follows from Lemma 2 that she adopts worldview $\alpha_1 = \bar{\alpha}_1$ obtaining first-stage payoff $(1 - \lambda_1)U_1(\bar{\alpha}_1, a) + \lambda_1 U_1(\alpha_1^0, a)$. Agent 1's responsibility to choose an alternative prevents her from further reducing her worldview: she knows that if she were to choose a worldview α_1 below $\bar{\alpha}_1$, then she could not prevent her future self from maintaining the status quo b, which she currently finds significantly inferior to a. Agent 1 therefore cannot "psychologically free-ride," by letting other agents worry about the problem on her behalf. Each pragmatist $i = 2, 3$, having no power over the choice of the alternative, simply seeks to maximize their first-stage payoff $(1 - \lambda_i)U_i(\cdot, a) + \lambda_i U_i(\alpha_i^0, a)$, and therefore chooses worldview $\alpha_i = 1$, since $v_i(a) \leq u_i(a)$. The inevitability of the action being taken forces pragmatists to accept it.

Pro-action pragmatist as dictator Suppose the voting rule is now $\mathcal{D} = \{C \subseteq N \colon C \ni 2\}$, so that pragmatist 2 the dictator. It is straightforward that Conditions (D) and (NFR) are satisfied with $I^{a*} = \emptyset$. Since $\alpha_2^0 = 1 > \alpha_2^*$, dictator 2 maintains her worldview at 1 and enforces a in equilibrium (recall Lemma 1). In contrast to the previous case, idealist 1 bears no responsibility for implementing policy action a. As a result she can avoid the tension between her current and future selves. Her future self fully avoids the psychological discomfort caused by acknowledging the problem and is in complete denial ($\alpha_1 = 0$). Simultaneously, her current self is satisfied that the problem will be addressed. The absence of power is, in fact, beneficial to agent 1. Her payoff when agent 2 is dictator is $(1-\lambda_1)U_1(0, a) + \lambda_1 U_1(\alpha_1^0, a)$, allowing her to gain $(1-\lambda_1)\bar{\alpha}_1\big[v_1(a) - u_1(a)\big]$. This value represents the discomfort avoided by agent 1. In contrast, pragmatist 3 shifts to worldview 1 to be comfortable with the inevitable choice.

Anti-action agent as dictator The previous examples illustrate how a change in the decision rule can have a substantial impact on the distribution of worldviews and payoffs in society. Now suppose the voting rule is $\mathcal{D} = \{C \subseteq N \colon C \ni 3\}$, so that pragmatist 3 is the dictator. Condition (NFR) can be satisfied with $I^{a*} = \emptyset$, but Condition (D) fails since 3 belongs to neither I^a nor P^a. No action is thus the unique equilibrium outcome and all agents choose complete denial of the problem ($\alpha_i = 0$ for all i). All agents are able to avoid any psychological discomfort and payoffs are maximized. This example not only shows how payoffs change, but how the institution can affect polarization in society. With pragmatist 3 as dictator no action is taken, and, hence, everyone prefers to deny the problem. In contrast to the prior two examples, there exists no polarization of worldviews in equilibrium. The distribution (and polarization) of worldviews in society depend critically on the *initial worldviews* of winning coalitions. The same result would hold if idealist 1 was the dictator and anti-action as opposed to pro-action. We expand on the relationship between inaction and polarization in the next section.

We next consider the following example of majority voting with and without a veto to make the denialist-free-riding effect starker.

Majority Voting with and without a veto Suppose first that decision-making is by simple majority-voting, i.e., $\mathcal{D} = \{C \subseteq N : |C| \geq 2\}$. Condition (D) holds with $I^{a*} = \{1\}$, however Condition (NFR) fails because $(I^{a*} \cup P) \setminus \{1\} = \{2, 3\}$ is a majority coalition. Although condition (D) ensures that there is a decisive coalition $P \cup I^{a*} = \{1, 2\}$ that supports the action, the failure of (NFR) drives its member 1 out of doing so. Indeed, if both 1 and 2 voted for a in equilibrium, then pragmatist 3 would do the same (Lemma 1). This, in turn, would imply that pro-action idealist 1 could profitably deviate by relying on the majority coalition $\{2, 3\}$ to enforce the action, while setting her own worldview at $\alpha_1 = 0$. However, both conditions hold if 1 is given a veto. As before, $\{1, 2\}$ constitutes a decisive coalition, so that (D) is still satisfied with $I^{a*} = \{1\}$. Moreover, coalition $\{2, 3\}$ is no longer decisive, since it does not include the vetoer, ensuring that (NFR) now holds. It follows that policy action a is selected in equilibrium because being compelled to be part

of any decisive coalition, pro-action idealist 1 can no longer free ride on coalition $\{2,3\}$. She cannot ignore the inclination of her initial worldview, but sets $\alpha_1 = \bar{\alpha}_1$ to minimize the psychological discomfort.

6 Formation of Worldviews: Institutions and Reform Feasibility

With the equilibrium characterization in hand, we can now return to the questions we raised in the Introduction. In particular, what do Conditions (D) and (NFR) teach us about the political feasibility of necessary reforms in the face of intangible problems, as well as the impact of collective choice institutions on the endogenous formation of individual worldviews regarding such issues? Does decision-making power constitute an advantage for agents as it is typically the case? These are the questions addressed in turn in this section. It is convenient to use the following definition of enforceability.

Definition 1. *Policy a is* enforceable *under collective choice rule $\mathcal{D}$ if Conditions* (D) *and* (NFR) *hold for some $I^{a*} \subseteq I^a$.*

6.1 Collective-Choice Institutions and the Feasibility of Policy Action

Institutional tradeoffs. Suppose that policy action a constitutes an improvement from a social welfare perspective, and consider the problem faced by a "constitution maker" aiming to implement institutional reforms to facilitate the enactment of this action. From Proposition 1, her task comes down to enhancing its enforceability. Intuition from standard collective choice problems with exogenous worldviews suggests that an appropriate reform for this constitution maker could be to adopt a more *permissive* voting rule—or more formally, to adopt a new rule $\mathcal{D}$ that includes more decisive coalitions. In the case of quota rules, for example, this simply consists in reducing the quota. By increasing the number of decisive coalitions, she would increase the likelihood that at least one of them will favor the policy action. This is precisely what Condition (D) in Proposition 1 tells us: if the collection $\mathcal{D}$ is larger, then it is easier to find a coalition I^{a*} such that $I^{a*} \cup \{j \in P\colon \alpha_j^0 \geq \alpha_j^*\}$

is decisive.

However, it follows from the analysis above that this seemingly implacable logic does not necessarily apply in the case of intangible problems where agents' worldviews are endogenously determined. The difference lies in Condition (NFR). If the voting rule is made more permissive, i.e., if $\mathcal{D}$ is expanded, then it becomes more likely that there be an idealist i for whom the coalition $(I^{a*} \cup P) \setminus i$ belongs to $\mathcal{D}$, thus violating enforceability. Intuitively, such a rule change would increase the idealists' opportunities to free ride on others' acknowledgement of the problem.

Thus, enforceability reveals *a tension between the need to forge a pro-action coalition and the need to limit the idealists' free-riding incentives.* As established below, this tension can prove to be unsolvable in the case of some rules that are empirically very relevant.

Quota rules and anonymity. A property often regarded as desirable for a voting rule to have is *anonymity* (e.g., Arrow, 1951). It requires that the identity of the agents does not matter in the aggregation of their preferences; more formally, a permutation of the agents' preferences over a and b does not change the outcome of the vote in the second stage of the game. Combined with the properties of properness and monotonicity, this property actually requires that $\mathcal{D}$ be a quota rule, so that there exists $q \in (n/2, n]$ such that $\mathcal{D} = \big\{C \subseteq N \colon |C| \geq q\big\}$. Owing to their common use in real-world applications, these rules have received considerable attention in the political-economy literature.

What makes the anonymity property of quota rules desirable is that agents are treated equally. Another implication of this property, however, is that agents are easily interchangeable when it comes to forming decisive coalitions. This substitutability fosters free-riding incentives, thereby making it challenging to achieve enforceability of policy action a. To see this, suppose that the coalition $\{j \in P \colon \alpha_j^0 \geq \alpha_j^*\}$ is not decisive, meaning that some pro-action idealists are required to meet the quota and form a decisive coalition (and satisfy Condition (D)). Anonymity of quota rules however induces any such idealist to "free ride" on the low-worldview pragmatists in $\{j \in P \colon \alpha_j^0 < \alpha_j^*\}$ who will inevitably vote for a in the voting stage (Lemma 1), making it impossible to satisfy Condition (NFR). By

Proposition 1, this in turn implies that alternative a cannot be sustained in a nontrivial equilibrium. It is worth noting that this logic applies *regardless of the value of the quota*: although reducing q makes the voting rule more permissive, this effect is offset by the increased opportunities for free riding that a lower q also creates.

Corollary 1. *Suppose* $\mathcal{D} = \{C \subseteq N : |C| \geq q\}$*, for some* $q > n/2$*. If* $\left|\{j \in P : \alpha_j^0 \geq \alpha_j^*\}\right| < q$ *and* $\{j \in P : \alpha_j^0 < \alpha_j^*\} \neq \emptyset$*, then policy action* a *cannot be sustained in a nontrivial equilibrium.*

As a stark example to illustrate Corollary 1, suppose that the group comprises $n-1$ idealists who all initially strongly endorse policy action (i.e., $\alpha_i^0 \approx 1$, for all $i \in I$), and a single pragmatist who weakly opposes it (i.e., $\alpha_j^0 < \alpha_j^*$ but $\alpha_j^0 \approx \alpha_j^*$, for $j \in P$). Thus, the action initially has overwhelming support. It follows however from the corollary above that, despite strong support from the group and regardless of the quota, it cannot be sustained in a nontrivial equilibrium. In this sense, Corollary 1 shows that *endogenous worldviews are conducive to political failures.*

Veto rights and unanimity rule. Corollary 1 suggests that the political sustainability of the policy action typically requires a violation of the anonymity criterion. As seen in the simple example in Section 5, one way to avoid free-riding incentives is to grant veto power to those agents likely to have such incentives. Indeed, if veto power is granted to a sufficient number of pro-action idealists, then Condition (NFR) of Definition 1 is trivially satisfied. All that remains to ensure enforceability is the decisiveness condition that is typically found in collective-choice problems, i.e., Condition (D). We summarize this in the corollary below.

Corollary 2. *Suppose there exists a set of pro-action action idealists* I^{a*} *such that* $I^{a*} \subseteq \bigcap \mathcal{D}$*, i.e, all members of* I^{a*} *have a veto, then, policy action* a *is sustainable in a nontrivial equilibrium if* I^{a*} *satisfies Condition* (D)*.*

While allowing more veto rights can relax the (NFR) constraint, one must be careful as illustrated by the case of unanimity. Unanimity makes each voter pivotal and thus relaxes

(NFR) entirely. However, it is clear that unanimity suffers from the typical challenge of satisfying decisiveness (D), as it requires each idealist and each pragmatist to be pro-action based on initial worldviews. We can conclude from this discussion that in order to solve intangible problems, collective-choice rules must strike a balance between reducing anonymity to avoid free riding and allowing sufficient flexibility to permit decisiveness.

Impact of mindset flexibility on institutional tradeoffs. Our application of the Bernheim et. al (2021) framework connects our work to the concept of "avoidance coping" that has been established in the psychology literature. As mentioned previously, Lazarus and Folkman (1984) divide coping strategies into (i) problem-focused coping, and (ii) emotion-focused coping. In our model, agents with high mindset flexibility (low λ_i) are able to easily adjust their world view. Since agents have a preference for ignoring problems and taking no action, mindset flexibility allows them to deny the existence of the problem, i.e., employ avoidance coping. This also implies that mindset flexible agents are less likely to employ problem-focused coping, hence take no action.

Following the previous logic, it is unsurprising that the policy action is more likely to be enforceable in societies that are mindset-*in*flexible. Indeed, highly mindset-inflexible agents are more inclined to supporting the action, as they are less inclined to engage in denial of the problem — the pro-action cutoffs α_i^* and α_i^{**} in Proposition 0 strictly decrease with the λ_i's. It follows that the decisiveness condition is more easily satisfied, without affecting the no-free-riding condition, when agents are highly mindset-inflexible.

Corollary 3. *Let $\lambda = (\lambda_1, \ldots, \lambda_n)$ and $\lambda' = (\lambda'_1, \ldots, \lambda'_n)$ be two mindset-inflexibility profiles such that $\lambda_i \leq \lambda'_i$, for all $i \in N$. All else being equal, if a is enforceable under λ, then it is enforceable under λ'.*

6.2 How Collective Decision Making Shapes Denial Behavior

So far, our discussion has mainly focused on the collective choice of an alternative in the voting stage. We now shift our focus to the individual choices of worldviews in the model's initial stage. It is typical in political economy to examine how the distribution of

policy preferences affects collective choices, but our model also allows us to study how the collective decision-making process influences the distribution of worldviews in equilibrium.

Impact of collective decision making on individual choices. A comparison of Propositions 0 and 1 allows us to analyze how the collective choice process affects agents' worldviews. Does it lead them to acknowledge the problem, or does it push them further into denial?

As we saw above, the enforceability of policy action requires a delicate balance between the decisiveness and no-free-riding conditions. When the collective-choice rule fails to achieve this balance and at least one of these two conditions is violated, pro-action agents—i.e., members of $I^a \cup P^a$ — undergo a radical shift in perspective compared to the worldviews they would have adopted had they been the sole decision-makers. The collective choice process drives them into complete denial of the problem and to vote for inaction. As for anti-action agents, they continue to deny the problem and support inaction, just as they would have on their own.

When the policy action is enforceable, the situation is different. Pro-action agents within the enforcing coalition $I^{a*} \cup P^a$ remain unaffected by the collective choice process: each $i \in I^{a*}$ adopts her cutoff worldview $\bar{\alpha}_i$, and each $j \in P^a$ adopts a worldview of 1; they still vote for the policy action. The agents whose worldviews differ in a collective-choice setting compared to when deciding individually are the pro-action idealists in $I^a \setminus I^{a*}$ and the anti-action pragmatists — i.e., the members of $P \setminus P^a$. While the collective choice process leads the former into complete denial as in the previous case, the latter undergo a drastic shift from denial to full acknowledgment of the problem. Unable to prevent the implementation of policy action by the collective, they are better off convincing themselves of its necessity. This in turn lead them to vote in favor of policy action.

We record these observations as:

Corollary 4. *Fix agent $i \in N$ and a voting rule $\mathcal{D}$. Then agent i's worldview choice α_i in every non-trivial equilibrium satisfies:*
(i) $\alpha_i > \alpha_i^d$ if and only if a is enforceable under $\mathcal{D}$ and $i \in P \setminus P^a$; and

(ii) $\alpha_i < \alpha_i^d$ *if and only if either* a *is enforceable under* $\mathcal{D}$ *and* $i \in I^a \setminus I^{a*}$*; or* a *is not enforceable under* $\mathcal{D}$ *and* $i \in I^a \cup P^a$*.*
Her voting behavior in such an equilibrium differs from her behavior as a dictator if and only if $\alpha_i \neq \alpha_i^d$*.*

We conclude from Corollary 4 that, unless an agent is a pragmatist who expects the group to take action against the problem, the collective decision-making process generally drives individuals toward denialist behavior.

In fact, it may be in the interest of proactive idealists to refuse or relinquish decision-making powers that would not allow such behavior. In typical political-economic settings where agents' policy preferences are fixed, it is evident that more political power always benefits any agent by enabling her to influence collective decisions in her favored direction. As illustrated in the simple example of dictators in Section 5, however, this is not the case when agents face policy problems that prompt them towards self-deception. For pro-action idealists, political decisiveness comes with heavy responsibilities: despite the discomfort it brings, they can no longer evade the reality of the problem; they must be part of the decisive coalition that enforces the action. Their situation embodies the famous adage "with great power comes great responsibility."

Polarizing action and cohesive inaction. An implication of Proposition 1 is that *the agents' equilibrium worldviews tend to be highly polarized when the policy action* a *prevails, whereas in sharp contrast, they perfectly align when the inaction status quo* b *is maintained.* In other words, while the action may be approved, it typically remains as a point of contention among the members of the group; whereas there is ultimately no disagreement when the status quo is preserved.

When agents expect the reform to be implemented, pragmatists raise their worldviews as much as possible (Lemma 1), while idealists lower theirs to the greatest extent (Lemma 2). The former benefit from convincing themselves that the problem is being addressed, whereas the latter seek to minimize their utility losses by worrying less about it. The only agents who refrain from adopting extreme worldviews in $\{0, 1\}$ are the idealists in

the critical coalition I^{a*}, who support the action but cannot free ride on others' worldview choices to ensure its adoption. As a result, the conflict of interest between supporters and opponents of the policy is moderated only by the members of I^{a*}. When this coalition is empty, worldviews become fully polarized between supporters and opponents of the action.

To formalize these observations, define the measure of *worldview polarization between supporters and opponents of the action* in any non-trivial equilibrium as

$$\Pi \equiv \frac{1}{|A||B|} \sum_{i \in A} \sum_{j \in B} |\alpha_i - \alpha_j| \in [0, 1] \ ,$$

where A denotes the coalition of agents who vote for a, and B represents those who vote for b in equilibrium. Naturally, we set $\Pi \equiv 0$ if either A or B is empty.

Proposition 2. *Suppose $P \neq \emptyset$. Then, there exists $k \in \left[0, 1/|P \cup I^a|\right]$ such that the following holds in any nontrivial equilibrium: $\Pi \geq 1 - k|I^{a*}| > 0$ if a is enforceable; and $\Pi = 0$ if a is not enforceable. In particular, $\Pi = 0$ if and only if $\alpha_i = 0$, for all $i \in N$.*

Proposition 2 is consistent with the literature on morality policy. Mooney (1999) points out that morality policies are highly salient, involve speaking out and active political participation. Our analysis reveals that this salience and activity results when the policy action to address this problem is being taken, not when it is being ignored.

Real-world examples abound of situations where policy action was accompanied with a division between staunch supporters and outright deniers, and inaction with widespread denial. Climate change is a prime example because its effects are long-term and often geographically distant. From 1979 (when the World Climate Conference recognized the dangers of climate change) to the early 2000s, many countries, particularly those with strong fossil fuel interests, displayed widespread denial or at least indifference toward the issue. In the United States, for example, both the Reagan administration (1981-1989) and the George H.W. Bush administration (1989-1993) were notably resistant to binding international agreements and largely dismissed the urgency of the problem. The Clinton administration (1993-2001), which advocated for climate action (e.g., the Kyoto Protocol), faced significant opposition in Congress, and many Americans showed limited concern. During this period, there was little polarization; climate change was not yet a deeply

partisan issue, and the general public maintained a more indifferent or denying stance, with little conflict. However, with the Obama administration pushing for more ambitious climate action, particularly the Clean Power Plan (2015), climate change became highly contentious. In particular, the Paris Agreement (2015), signed by the President, was accompanied by intense polarization: supporters of climate reform (mainly Democrats and environmentalists) recognized the scientific consensus and advocated for policy action, while opponents (often Republicans and the fossil fuel industry) denied or minimized the issue.

Similarly, mass displacements due to war or economic collapse are often perceived as distant problems that do not immediately affect domestic populations. Public opinion tends to be divided between those in favor of humanitarian aid who fully recognize the crisis, and opponents who often downplay its severity or legitimacy (e.g., claiming that refugees are "economic migrants" rather than victims of war). The European migrant crisis (2015) illustrated this, with some nations accepting refugees and others denying the humanitarian urgency. For instance, Hungary largely rejected refugees, with Prime Minister Viktor Orbán calling the crisis a "mass invasion" and erecting border fences. Public opinion in Hungary was overwhelmingly opposed to accepting refugees, with little domestic polarization. In contrast, Germany's Chancellor Angela Merkel implemented the "Wir schaffen das" ("We can do this") policy, welcoming over one million refugees. As in our model, this policy sparked strong polarization: supporters saw it as a moral and legal obligation, while opponents warned of security risks and cultural tensions (e.g., the rise of the far-right AfD party in Germany).

6.3 Psychological Coping Strategies and Mindset Flexibility

Our model offers a collective decision-making perspective on Lazarus and Folkman's (1984) influential theory of coping strategies in psychology, which distinguishes between problem-focused and emotion-focused coping. The former strategy involves actively addressing a stressor to reduce or eliminate it, while the latter aims to manage the emotional distress associated with the stressor. In particular, they identify "avoidance coping" as a subtype

of emotion-focused coping, which involves by ignoring, denying, or distancing oneself from the problem. While their theory focuses on *individual* decision-making, our equilibrium analysis extends these insights to *collective* approaches to coping with ignorable problems.

A key implication of our results is that *collective decision-making processes tend to drive groups toward avoidance coping rather than problem-focused coping when dealing with ignorable problems,* and to a greater extent than any individual group member would do on their own.

A key factor influencing agents' incentives to engage in denialist behavior in our model is Bernheim et al.'s (2021) concept of mindset flexibility. Our application of this concept reveals a connection to Lazarus and Folkman's (1984) theory. In our model, agents with high mindset flexibility (low λ_i) are more inclined to adjust their worldview. Given that agents have an inherent preference for ignoring problems and taking no action, mindset flexibility enables them to deny the problem's existence. Consequently, highly mindset-flexible agents are also less likely to employ problem-focused coping and, therefore, take no action.

Following the previous logic, it is then unsurprising that policy action is more likely to be enforceable in societies characterized by mindset *in*flexibility. Indeed, highly mindset-inflexible agents are more inclined to support action, i.e., to adopt a problem-focused strategy in Lazarus and Folkman's language. This is reflected in the fact that the pro-action cutoffs α_i^* and α_i^{**} in Proposition 0 strictly decrease with λ_i. As a result, the decisiveness condition is more easily met without compromising the no-free-riding condition when agents exhibit high mindset inflexibility.

Corollary 5. *Let $\lambda = (\lambda_1, \ldots, \lambda_n)$ and $\lambda' = (\lambda'_1, \ldots, \lambda'_n)$ be two mindset-inflexibility profiles such that $\lambda_i \leq \lambda'_i$, for all $i \in N$. All else being equal, if a is enforceable under λ, then it is enforceable under λ'.*

7 Concluding Remarks

This paper examines the political feasibility of policy actions addressing "ignorable problems," i.e., problems that do not directly and immediately affect individuals' utilities. The analysis of such problems raises a fundamental question: the interplay between *collective* policy choices and the *individual* formation of worldviews regarding the existence of policy problems. In particular, we show that when individual worldviews are endogenous, the political feasibility of a policy no longer depends solely on the existence of a decisive coalition of supporters (decisiveness). It also requires a second condition, no-free-riding, ensuring that idealists within such a coalition cannot delegate the burden of acknowledging the problem to others, while ignoring it themselves.

Its simplicity notwithstanding, our model has significant implications for the political sustainability of action on morality policies and denial behavior in collective decision-making contexts. In particular, it highlights a problematic tension between the conditions of decisiveness and no-free-riding, which can prove insurmountable under the empirically relevant anonymous voting rules. It also shows that collective decision-making makes individuals more prone to denying the problem than when they face it alone. Moreover, there is a striking asymmetry between equilibria where the policy action is implemented and those where the inaction status quo persists: in the former, strong polarization of worldviews emerges between supporters and opponents of the policy, while in the latter, unanimous denial of the problem prevails.

References

Alesina, A., and A. Drazen (1991) "Why Are Stabilizations Delayed?" *American Economic Review* 81, 1170-1188.

Andreoni, J. (1989) "Giving with Impure Altruism: Applications to Charity and Ricardian Equivalence," *Journal of Political Economy* 97, no. 6 (1989): 1447-1458.

Anesi, V., and T.R. Bowen (2021) "Policy Experimentation in Committees: A Case against Veto Rights under Redistributive Constraints," *American Economic Journal: Microeconomics* 13, 124-162.

Anesi, V., and D.J. Seidmann (2015) "Bargaining in Standing Committees with an Endogenous Default," *Review of Economic Studies* 82, 825-867.

Arrow, K. J. (1951) *Social Choice and Individual Values.* New Haven, CT: Yale University Press.

Austen-Smith, D., and J.S. Banks (2000) *Positive Political Theory I: Collective Preference.* University of Michigan Press.

Bénabou, R. (2013) "Groupthink: Collective Delusions in Organizations and Markets," *Review of Economic Studies* 80, 429-462.

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

Bénabou, R., and J. Tirole (2006) "Incentives and Prosocial Behavior," *American Economic Review* 96, 1652-1678.

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

Bernheim, B.D., L. Braghieri, A. Martínez-Marquina, and D. Zuckerman (2021) "A Theory of Chosen Preferences," *American Economic Review* 111, 720-754.

Bowen, T.R., D. Dmitriev, and S. Galperti (2023) "Learning from Shared News: When Abundant Information Leads to Belief Polarization," *Quarterly Journal of Economics* 138, 955-1000.

Clayton, S., and B.T. Karazsia (2020) "Development and Validation of a Measure of Climate Change Anxiety," *Journal of Environmental Psychology* 69, 101434.

Coate, S., and S. Morris (1999) "Policy Persistence," *American Economic Review* 89, 1327-1336.

Cohen, S. (2013) *States of Denial: Knowing about Atrocities and Suffering.* John Wiley & Sons.

DeMarzo, P.M. (1992) "Coalitions, Leadership, and Social Norms: The Power of Suggestions in Games," *Games and Economic Behavior* 4, 72-100.

Dunn, E.W., L.B. Aknin, and M.I. Norton (2008) "Spending Money on Others Promotes Happiness," *Science* 319, 1687-1688.

Dziuda, W., and A. Loeper (2016) "Dynamic Collective Choice with Endogenous Status Quo," *Journal of Political Economy* 124, 1148-1186.

Ekberg, K., B. Forchtner, M. Hultman, and K.M. Jylhä (2022) *Climate Obstruction: How Denial, Delay and Inaction are Heating the Planet.* Taylor & Francis.

Fernandez, R., and D. Rodrik (1991) "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty," *American Economic Review* 81, 1146-1155.

Hale, T. (2024) *Long Problems: Climate Change and the Challenge of Governing Across Time.* Princeton University Press.

Howitt, P., and R. Wintrobe (1995) "The Political Economy of Inaction," *Journal of Public Economics* 56, 329-353.

Izzo, F., G.J. Martin, and S. Callander (2023) "Ideological Competition," *American Journal of Political Science* 67, 687-700.

James, W. (1890) *Principles of Psychology.* Dover Publications.

Kals, E., and J. Maes (2002) "Sustainable Development and Emotions." In: Schmuck, P., Schultz, W.P. (eds) *Psychology of Sustainable Development.* Springer, Boston, MA.

Lazarus, R.S., and S. Folkman (1984) *Stress, Appraisal, and Coping.* New York: Springer.

Levy, G., R. Razin, and A. Young (2022) "Mispecified Politics and the Recurrence of Populism," *American Economic Review* 112, 928-962.

Levy, G., and R. Razin (2025) "Political Social-Learning: Short-Term Memory and Cycles of Polarization," *American Economic Review* 115, 635-659.

Nietzsche, F. (1968) *Beyond Good and Evil: Prelude to a Philosophy of the Future.* In W. Kaufman (Ed. and Trans.), *Basic Writings of Nietzsche* (pp. 118-435). New York: The Modern Library. (Original work published 1886.)

Norgaard, K. M. (2011) *Living in Denial: Climate Change, Emotions, and Everyday Life.* MIT Press.

Nunnari, S., and J. Zápal (forthcoming) "A Model of Focusing in Political Choice," *Journal of Politics.*

Olsen, K., and K. Aas (2019) "Public Pension Schemes and Retirement Planning: Evidence from a Norwegian Pension Reform," *Journal of Pension Economics and Finance* 18, 570-587.

Perego, J., and S. Yuksel (2022) "Media Competition and Social Disagreement," *Econometrica* 90, 223-265.

Strulovici, B. (2010) "Learning while Voting: Determinants of Collective Experimentation," *Econometrica* 78, 933-971.

Whitmarsh, L., and S. Capstick (2018) "Perceptions of Climate Change." In *Psychology and Climate Change* (pp. 13-33). Academic Press.

Yuksel, S. (2022) "Specialized Learning and Political Polarization," *International Economic Review* 63, 457-474.

Zimmermann, F. (2020) "The Dynamics of Motivated Beliefs," *American Economic Review* 110, 337-363.

Appendix

A Proof of Proposition 1

We begin with the following two results.

Lemma A1. *In any equilibrium the set C of idealists i who adopt $\bar{\alpha}_i$ is empty if b is the winning alternative.*

Proof. Consider an equilibrium in which the status quo b is upheld, and suppose that C is nonempty. In this case, every i in C could profitably change her equilibrium worldview choice $\bar{\alpha}_i$ to 0. This would not have any impact on the outcome of the vote, but it would increase i's first-stage payoff by $(1-\lambda_i)\big[v_i(b) - U_i(\bar{\alpha}_i, b)\big] > 0$ — a contradiction. Hence, all the idealists adopt worldview 0 and then vote for the status quo b in any equilibrium where it is maintained. □

Lemma A2. *Fix a non-trivial equilibrium, and let C be the set of idealists i who adopt $\bar{\alpha}_i$ in this equilibrium. If C is nonempty, then C must satisfy:*

(i) $C \cup P^a$ is decisive;

(ii) for every $i \in C$, the coalition $(C \cup P) \setminus \{i\}$ is not decisive.

Proof. Consider a nontrivial equilibrium in which C is nonempty, so that by Lemma A1, a is the winning alternative. As the equilibrium worldviews α must constitute a strong Nash equilibrium of the worldview-choice game, C must satisfy Condition (D) in Proposition 1. Otherwise, the members of $\{j \in P\colon \alpha_j^0 < \alpha_j^*\}$ could profitably deviate by all adopting worldview 0, inducing all the members of the decisive coalition $\{j \in P\colon \alpha_j^0 < \alpha_j^*\} \cup (I \setminus C)$ to vote for (and enforce) b.

Moreover, C must also satisfy Condition (NFR). Indeed, we know from Lemma 1 that, like the members of C, all the pragmatists in P vote for a in such an equilibrium. If there were an i in C such that $(C \cup P) \setminus \{i\} \in \mathcal{D}$, then i could profitable deviate by changing her

equilibrium worldview choice to 0. She would thus improve her future self's utility without affecting the outcome of the vote. This would in turn increase her first-stage payoff by

$$(1-\lambda_i)\big[v_i(a)-U_i(\bar{\alpha}_i,a)\big]=(1-\lambda_i)\bar{\alpha}_i\big[v_i(a)-u_i(a)\big]>0$$

— a contradiction. □

We now return to the proof of the main proposition. We begin by showing that Conditions (D) and (NFR) hold for some $I^{a*}\subseteq I^a\equiv\{j\in I\colon \alpha_j^0\geq\alpha_j^{**}\}$, then it is possible to construct an equilibrium in which policy a is implemented. Consider the strategy profile σ that prescribes all the pragmatists to adopt worldview 1, all the members j of I^{a*} to adopt worldview $\bar{\alpha}_j$, and all the other idealists (if any) to adopt worldview 0, in the worldview-choice stage. In the voting stage, agents are prescribed to vote of a if and only if their chosen worldview is greater than or equal to their $\bar{\alpha}_i$. Thus, on the equilibrium path, all the pragmatists and the members of C vote for a — which is therefore implemented — and the remaining idealists vote for b. It is readily checked that individual deviations from σ are unprofitable, so that it constitutes an equilibrium.

To see that equilibrium σ is nontrivial, consider coalitional deviations in the initial stage. Observe first that by Condition (D), the coalition $\{j\in P\colon \alpha_j^0\geq\alpha_j^*\}\cup I^{a*}\subseteq\{j\in P\colon \alpha_j^0\geq\alpha_j^*\}\cup\{j\in I\colon \alpha_j^0\geq\alpha_j^{**}\}$ is decisive and, consequently, no coalition can profitably deviate by enforcing the status quo b — recall that $\mathcal{D}$ is proper. Hence, any candidate deviation would have to lead to a still being the winning alternative. Given this constraint, no idealist outside I^{a*} would agree to a coalitional deviation: as $u_i(a)<v_i(a)$, $\alpha_i=0$ is the unique maximizer of $(1-\lambda_i)U_i(\cdot,a)+\lambda_iU_i(\alpha_i^0,a)$, for all $i\in I$. The same applies to all the pragmatists: as $v_j(a)<u_j(a)$, $\alpha_j^1=1$ is the unique maximizer of $(1-\lambda_j)U_j(\cdot,a)+\lambda_jU_j(\alpha_j^0,a)$, for all $j\in P$. As for the members i of I^{a*}, they cannot adopt worldviews below $\bar{\alpha}_i$: by Condition (NFR), if any of them would do so, then the status quo b would be maintained, thus making them all worse off. As their payoffs $(1-\lambda_i)U_i(\cdot,a)+\lambda_iU_i(\alpha_i^0,a)$ are strictly decreasing, they cannot do better than all choosing $\alpha_i=\bar{\alpha}_i$.

To complete the proof of the first part of Proposition 1, it remains to show that b cannot

be a non-trivial-equilibrium policy outcome if some I^{a*} satisfies Conditions (D) and (NFR). By Condition (D), all the members of the decisive coalition $\{j \in P\colon \alpha_j^0 \geq \alpha_j^*\} \cup I^{a*}$ initially prefer a to b. If b were the winning alternative, then this coalition could make all of its members strictly better off by adopting different worldviews (1 for its pragmatist members, and $\bar{\alpha}_i$ for its idealist members), thus violating nontriviality.

Equilibrium worldview choices follow directly from Lemmas 1 and 2. To complete the proof of the proposition, therefore, it remains to establish that a cannot be a non-trivial equilibrium policy outcome if there is no I^{a*} that satisfies Conditions (D) and (NFR). We prove this by contrapositive. Consider any nontrivial equilibrium in which alternative a is implemented. By definition of nontriviality, there must be a decisive coalition of agents who, initially, prefer alternative a to b; i.e., $\{j \in P\colon \alpha_j^0 \geq \alpha_j^*\} \cup \{j \in I\colon \alpha_j^0 \geq \alpha_j^{**}\} \in \mathcal{D}$. If $\{j \in P\colon \alpha_j^0 \geq \alpha_j^*\} \in \mathcal{D}$, then Conditions (D) and (NFR) trivially hold for $I^{a*} = \emptyset$. If $\{j \in P\colon \alpha_j^0 \geq \alpha_j^*\} \notin \mathcal{D}$, then some idealists must vote for a in equilibrium. By Lemma A2, this implies that there is a coalition I^{a*} that satisfies Conditions (D) and (NFR).

B Proof of Proposition 2

Let $m^a \in \left[0, |I^a|\right]$ denote the size of the largest coalition $I^{a*} \in 2^{I^a}$ that satisfies conditions (D) and (NFR) in Proposition 1. Then, define $k \equiv \left(|P| + m^a\right)^{-1}$. Observe that k is well-defined since P is assumed to be nonempty.

Suppose first that a is enforceable, so conditions (D) and (NFR) hold for some coalition of idealists I^{a*}. From Proposition 1, policy action a is the only nontrivial equilibrium; and $A = P \cup I^{a*}$ and $B = I \setminus I^{a*}$ are the coalitions of agents who vote for a and for b in equilibrium, respectively. In addition, $\alpha_i = 1$ for all $i \in P$ (Lemma 1), and $\alpha_j = 0$ for all

$j \in I \setminus I^{a*}$ (Lemma 2). Therefore, we have

$$\begin{aligned}
\Pi &\equiv \frac{1}{|A||B|} \sum_{i \in A} \sum_{j \in B} |\alpha_i - \alpha_j| \\
&= \frac{1}{\big[|P| + |I^{a*}|\big]|I \setminus I^{a*}|} \left[\sum_{i \in P} \sum_{j \in I \setminus I^{a*}} |\alpha_i - \alpha_j| + \sum_{i \in I^{a*}} \sum_{j \in I \setminus I^{a*}} |\alpha_i - \alpha_j| \right] \\
&\geq \frac{1}{\big[|P| + |I^{a*}|\big]|I \setminus I^{a*}|} |P||I \setminus I^{a*}| \\
&= 1 - \frac{|I^{a*}|}{|P| + |I^{a*}|} \geq 1 - k|I^{a*}| > 0 \ ,
\end{aligned}$$

as desired. (The last inequality follows directly from the definition of k.)

Now, suppose that a is not enforceable. Then, it follows from Proposition 1 that $A = \emptyset$ and $\alpha_i = 0$, for $i \in N$. This in turn implies that $\Pi = 0$, proving the proposition.

C Micro-Founded Self-Deception

As mentioned in footnote 2, we did not impose any restrictions on the agents' ability to self-deceive, in order to present the main intuitions of our model in the most parsimonious way. However, it is possible to incorporate more complex structures to describe explicitly to self-deception process, as does Bénabou (2013) for example. Following the latter's specification precisely, but adapting the terminology and notation to our model, suppose now that agents receive a fully informative signal $\sigma \in \{0, 1\}$ regarding the existence of the problem, where $\sigma = 1$ indicates that it is real. Upon observing $\sigma = 1$, each agent i decides (consciously or unconsciously) how much attention to allocate to the information, how to interpret it, or whether to retain or disregard it. Formally, the individual encodes a signal $\hat{\sigma}_i$ into memory, where $\hat{\sigma}_i = 1$ means acceptance of the facts, whereas $\hat{\sigma}_i = 0$ means denial, censorship, or rationalization. Her "awareness strategy" is then defined as $\mathsf{a}_i \equiv \Pr\{\hat{\sigma}_i = 1 \mid \sigma = 1\}$. It is assumed for simplicity that agents acknowledge good news $\sigma = 0$ with probability one.

Agent i initially believes that she will receive $\sigma = 1$ with probability α_i^0. Bénabou (2013) then posits that agent i's posterior belief/worldview α_i results from Bayes' updating,

i.e., $\Pr\{\sigma = 1 \mid \hat{\sigma}_i = 1\} = 1$ and

$$\Pr\{\sigma = 1 \mid \hat{\sigma}_i = 0\} = \frac{\alpha_i^0(1 - \mathsf{a}_i)}{1 - \mathsf{a}_i\alpha_i^0} \ ,$$

This comes down to assuming that she is constrained to choose her worldview α_i within her "self-deception set"

$$A_i \equiv [0, \alpha_i^0] \cup \{1\} \ .$$

Under this new specification, our equilibrium characterization remains exactly the same as previously. To see this, observe that among the possible optimal worldview choices of agent i (i.e, 0, 1, or $\bar{\alpha}_i$), 0 and 1 still belong to her self-deception set. As for $\bar{\alpha}_i$, it is only optimal for idealist i if $\alpha_i^{**} \leq \alpha_i^0$, in which case $\bar{\alpha}_i$ also belongs to A_i since $\bar{\alpha}_i < \alpha_i^{**}$.