

Status-Quo Losses and Constitutional Bargaining in Democratic Transitions

Taoufik Rajhi

taoufik.rajhi@univ-poitiers.fr

Université de Poitiers

Research Article

Keywords: democratic transitions, constitutional stability, elite resistance, loss aversion, reference-dependent preferences, transitional compensation

Posted Date: October 15th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11028163/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Status-Quo Losses and Constitutional Bargaining in Democratic Transitions

September 2026

Abstract

When does the timing and legal form of the concessions a new democracy makes to the elite it replaces matter beyond their present value? This paper studies an entitlement-reference model of elite resistance. A constitutional framework is in force while the incumbent elite can still overturn it, and that elite judges the transition against the position the prior order guaranteed it, weighting losses more heavily than gains. It may therefore resist a transition that serves its own present-value interest. A concession is worth more while the recipient remains below its inherited benchmark than once it has recovered, so the effectiveness of an otherwise identical concession jumps at the benchmark. That discontinuity, not the advantage of early concessions, separates the mechanism from accounts based on limited commitment, illiquidity, or short horizons. Because no government can commit to pay an elite whose veto has lapsed, credible concessions must be vested at the transition or entrenched in the constitution; and one concentrated in the transition period relaxes the elite's participation constraint by more than a recurring privilege of equal present value. An optimal settlement therefore concedes early, giving transitional bargains a reason to rely on one-off protections rather than permanent elite privileges.

Keywords: democratic transitions; constitutional stability; elite resistance; loss aversion; reference-dependent preferences; transitional compensation.

JEL Classification: D72, D74, D91, O43, P48.

1 Introduction

Democratic transitions frequently fail to consolidate. The recent comparative-politics literature documents a sustained episode of autocratisation spanning both consolidated democracies and fragile transitions (Bermeo, 2016; Lührmann and Lindberg, 2019; Haggard and Kaufman, 2021). A recurrent configuration is one in which a transition, set in motion by mass mobilisation or by the collapse of an incumbent, produces a constitutional framework that is in force before the elite that controlled the previous regime has accepted it. During that interval the incumbent elite typically retains coercive, economic, and bureaucratic capacity to overturn the new order, and settlements are negotiated: amnesties are legislated, transfers are paid, protections are written into the constitution. This paper asks a question about those settlements. When does their timing and form matter beyond their present value?

The mechanism through which timing can matter is behavioral. A democratic transition redistributes elite-controlled rents—patronage networks, procurement access, and regulatory protection of connected firms—and so imposes a short-run consumption loss on the elite before the democratic dividend materialises. If the elite evaluates that loss against the position the prior order guaranteed it, and if losses relative to that benchmark are weighted more heavily than equal-sized gains, the transition trough is behaviorally overweighted, and a sufficiently loss-averse elite resists. Throughout, *resistance* denotes the elite’s decision to withhold accommodation from a transition already in force and thereby to restore authoritarian rule; *reversal* and *backsliding* are reserved for observed historical episodes in which a transition ended.

Two features of the timing deserve emphasis, because the paper’s claims about early and late concessions are meaningless without them. First, the decision date is defined explicitly (Definition 1): it is the first period in which a democratic framework is in force while the elite’s capacity to overturn it is intact. Concessions are early or late relative to that window, not relative to the uprising that preceded it, and Section 9 dates the window for each case it discusses. Second, the elite’s choice is not artificially restricted to a single date. If the veto may be exercised in any period, it is exercised only during the transition trough, because from the first period of growth onward the elite is above its benchmark under either regime, the loss weight no longer applies, and accommodation dominates in level and in growth (Lemma 1). The model therefore has a window of vulnerability that coincides with the trough. Where a restoration could recover rents rather than merely return the elite to its counterfactual, that window widens, but the motive for late resistance is then a matter of present value and not of loss aversion (Remark 2). A breakdown long after the transition accordingly requires either an adapted benchmark and a new shock or actors other than the elite that accommodated, and the paper does not claim to account for events of that kind.

Within the window, the elite accommodates if and only if its *effective loss aversion* does not exceed the discounted democratic dividend measured per unit of transition cost. Effective loss aversion is the ratio of the marginal value of consumption just below the inherited benchmark to its marginal value just above, and it combines two preference parameters: the weight the elite attaches to gains and losses relative to intrinsic consumption value, and the extra weight it attaches to losses. The accommodation decision depends on preferences only through that ratio. Regime choices therefore identify the ratio and not its two components separately, which also means that experimental loss-aversion estimates, which are themselves ratios of marginal valuations, correspond to the composite rather than to the loss-aversion coefficient alone.

The behavioral content of the condition is confined to configurations in which the present-value comparison favours accommodation. Where the discounted dividend does not cover the transition cost, resistance is the prediction of any standard model and loss aversion adds nothing; where it does cover the cost, an elite that weighted losses and gains alike would accommodate, and resistance requires effective loss aversion above the threshold (Corollary 1). Within that region the mechanism implies something that no account working through present value alone implies. A concession raises the value of accommodation by more per unit when the recipient is below its benchmark than when it is above, so the effectiveness of an otherwise identical concession jumps at the benchmark, in the ratio just described (Proposition 1). This is the paper’s discriminating prediction, and it is deliberately not the prediction that early concessions work better: limited commitment, borrowing constraints, political survival risk, and short horizons all imply that, and Section 7.2 adds a commitment reason of its own. What none of them implies is a discontinuity at the recipient’s own inherited position. The claim is correspondingly modest: such a discontinuity rejects the class of smooth alternatives and is consistent with reference dependence, without identifying it, since institutional thresholds can produce jumps of their own—but at legal cut-offs rather than at inherited positions.

Two extensions test the robustness of the premise and of the environment. When the benchmark adjusts gradually towards realised consumption, the threshold is available in closed form, and adjustment works through two channels of opposite sign: it absorbs part of the trough into the benchmark, which favours accommodation, and it erodes the gains that the faster-growing democratic path records, which favours resistance. Under a mild and explicit condition on the weight of gain-loss utility, the second channel dominates at first, so moderate adjustment strengthens the mechanism rather than weakening it, and resistance is overturned only once adjustment becomes fast (Proposition 2). When a re-stored autocracy is itself liable to be displaced, and displacement would inflict a loss that persists, the threshold rises, and beyond a critical hazard loss aversion favours accommodation, because the same asymmetry that overweights the trough also overweights the

loss a failed restoration would bring (Proposition 3). Behavioral resistance is therefore concentrated where restoration is secure.

The constitutional analysis follows from the commitment structure. A government cannot commit to pay an elite whose veto has lapsed, so in equilibrium only concessions vested at the transition or entrenched in the constitution have value (Lemma 6). Three instruments remain: cash, bounded by fiscal and borrowing capacity and by the concentration of pre-transition rents; transitional protections such as amnesties and reconciliation procedures, whose value to the elite is concentrated in the transition period; and recurring constitutional privileges, which are credible in every period but distort policy for the life of the constitution. Their behavioral leverage differs. A transitional protection places its entire value in the interval in which the elite is below its benchmark, whereas a recurring privilege spreads value across a path of which only a fraction falls in that interval, so per unit of present value conceded the first relaxes the elite's participation constraint by more—about forty per cent more at the illustrative configuration, and more once the policy distortion of entrenchment is counted (Proposition 4). A designer acting for citizens therefore concedes in the trough—through whichever of cash and transitional protection citizens price more cheaply, since the two have identical leverage—and turns to recurring privilege last (Proposition 5); that ranking is global within the trough and local for entrenchment absent a monotonicity condition. And because citizens will not finance a settlement worth less to them than democracy itself, there are configurations in which no settlement in the admissible instrument set satisfies both sides and the transition fails (Proposition 6). That locates a bargaining frontier under the model's citizen objective; it is not a constitutional norm.

The contribution should be located precisely. It is not the resistance threshold, which is immediate once reference-dependent preferences are specified, and it is not the claim that loss-averse elites resist democracy, which is both too broad and dependent on the choice of benchmark. It is the constitutional instrument problem: when an outgoing elite evaluates constitutional change relative to inherited entitlements, a unit of concession delivered while those entitlements are in the loss domain carries strictly greater political leverage than the same present value embedded permanently in the constitution, which gives transitional bargains a reason to rely on one-off protections rather than on permanent elite privileges. The reference-crossing prediction is the empirical counterpart of the same asymmetry. The paper does not claim that resistance to democratic transitions is primarily behavioral, and the inherited benchmark is a primitive rather than a derivation: under an expectations-based reference the mechanism does not operate, and Section 4 states that dependence plainly. The empirical sections are illustrations: this is a theory paper, and Section 9 is explicit that its cases cannot validate the mechanism. The commitment-problem tradition (Acemoglu and Robinson, 2001) treats democratic collapse as the outcome of intertemporal bargaining under credible-commitment failure;

the political-agency tradition (Besley, 2006; Persson and Tabellini, 2009) treats it as selection and discipline of office-holders. Both abstract from the behavioral structure of elite preferences. Supplying that structure yields a formal reason why transitional settlements concentrate protection at the moment of transition rather than spreading it across the life of the constitution, which distinguishes pacted transitions (O'Donnell and Schmitter, 1986; Linz and Stepan, 1996) from the elite-biased constitutions of Albertus and Menaldo (2018): in this model the latter are the expensive instrument, chosen when the cheap ones are exhausted or unavailable.¹

The behavioral apparatus rests on reference-dependent preferences in the tradition of Tversky and Kahneman (1991); Kőszegi and Rabin (2006), and on the dynamic treatment of such preferences developed by Pagel (2017). The application of loss aversion to elite political choice connects to the behavioral-political-economy programme of Passarelli and Tabellini (2017) and Alesina and Passarelli (2019); the distinctive feature here is the entitlement reference, anchored in the political economy of lost rents rather than in a rational-expectations consumption forecast.

The remainder of the paper proceeds as follows. Section 2 states the motivating observations. Section 3 sets out the environment, the decision date, and the recurrent-veto result. Section 4 discusses the reference point. Section 5 derives the threshold, separates its present-value and behavioral regions, states the reference-crossing prediction, and characterises the threshold under gradual reference adjustment. Section 6 introduces fragile restoration. Section 7 studies the three instruments, their credibility, their behavioral leverage, the optimal settlement, and the conditions under which no settlement exists. Section 8 illustrates the magnitudes. Section 9 dates the decision window for three transitions, maps the primitives, and states the scope conditions. Section 10 concludes. Proofs are collected in Appendix A.

2 Motivating Facts

The figures in this section use two indices from version 16 of the Varieties of Democracy (V-Dem) dataset (V-Dem, 2026). The electoral democracy index, or polyarchy index, aggregates five components—freedom of association, clean elections, freedom of expression and alternative sources of information, elected officials, and suffrage—and ranges from 0 (least democratic) to 1 (most democratic). The liberal democracy index adds constraints on the executive, the rule of law, and the protection of individual liberties, and lies on the same unit interval. A polyarchy score above 0.5 is one of the conditions under which the Regimes of the World typology classifies a regime as an electoral democracy (Lührmann

¹The baseline operates without type heterogeneity, asymmetric information, signaling, or commitment problems, in order to isolate the benchmark mechanism. Extending it to privately informed elite types, with citizens inferring type from the observed regime choice, is left for separate work.

et al., 2018); it is used below as a descriptive threshold, not as a sharp classification.

Figure 1 traces Tunisia’s V-Dem scores from 1990 to 2024. Following the Ben Ali regime (polyarchy ≈ 0.19), the ouster of the president in January 2011 triggered a rapid democratisation: the polyarchy score reached 0.76 by 2014. A decade of competitive elections followed before President Saied’s self-coup in July 2021, after which the score fell to approximately 0.42 by 2024.

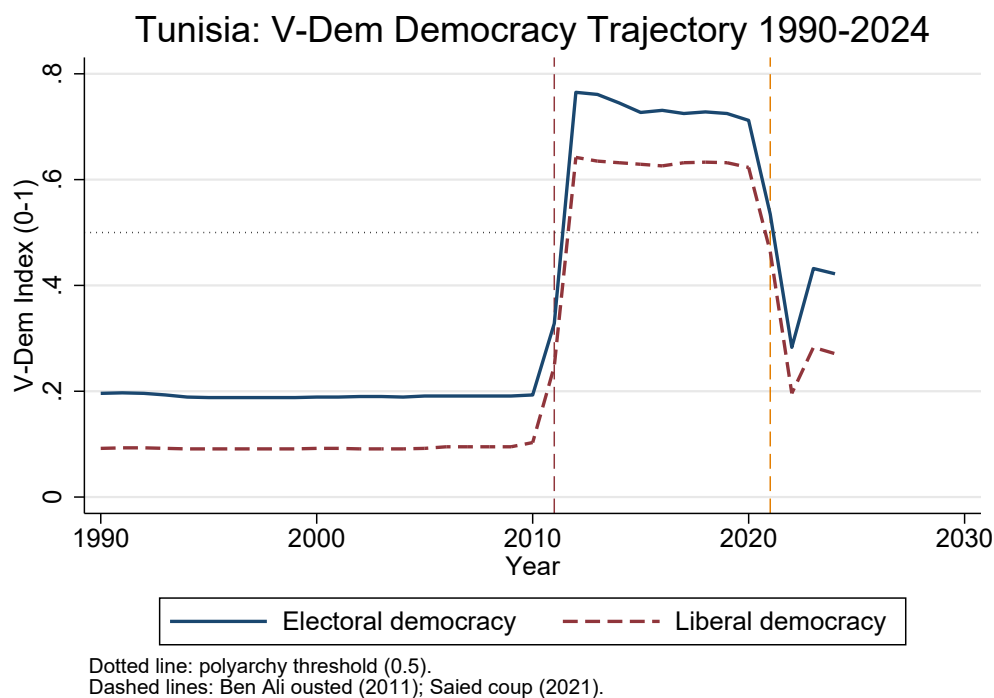


Figure 1: Tunisia: V-Dem polyarchy and liberal-democracy indices, 1990–2024. Dotted line: descriptive threshold (0.5). Vertical dashed lines: Ben Ali ousted (2011) and Saied self-coup (2021). Source: V-Dem v16.

Figure 2 presents Egypt’s trajectory. The polyarchy score peaked at approximately 0.35 in 2012 following the ouster of Mubarak— never crossing the descriptive threshold— before returning to authoritarian levels (≈ 0.13) following the military coup of July 2013.

Figure 3 compares the trajectories of all Arab League member states. Tunisia’s decade-long democratic episode is the case closest to the configuration studied in this paper; Egypt’s path is one of political disruption without democratisation.

Three features of democratic transitions motivate the model.

First, transitions typically impose front-loaded costs and back-loaded benefits. The dismantling of patronage structures, the renegotiation of procurement and licensing arrangements, and the institutional disruption of the early transition period precede the productivity gains that motivate democratisation. The relevant elite therefore faces a negative-then-positive consumption profile rather than a uniform improvement.

Second, a substantial political-economy literature suggests that elite rent dissipation

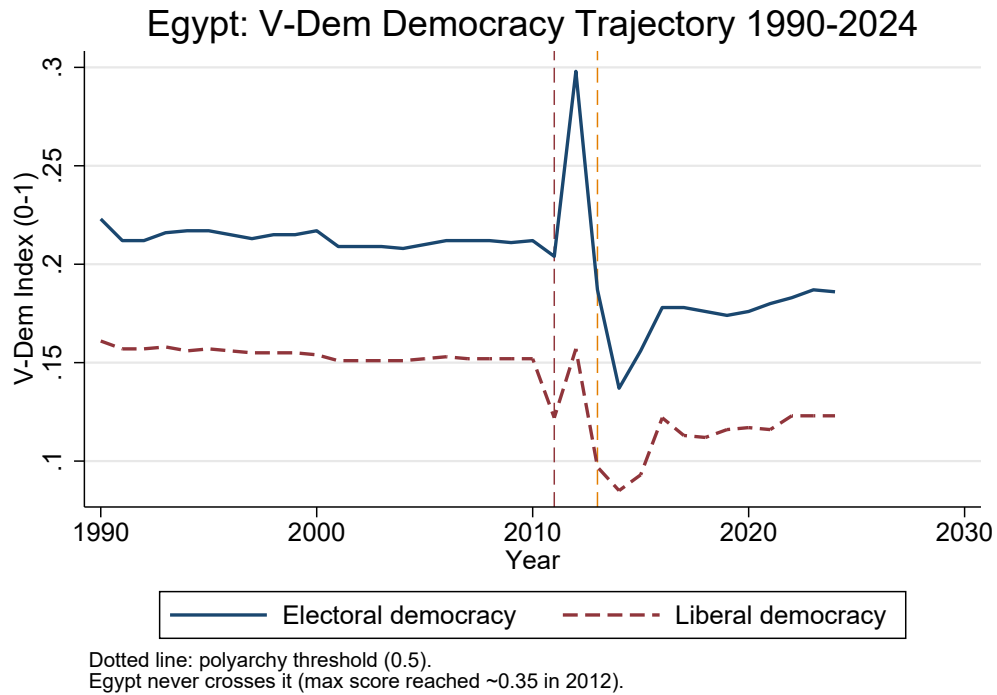


Figure 2: Egypt: V-Dem polyarchy and liberal-democracy indices, 1990–2024. Egypt never crosses the descriptive threshold of 0.5. Vertical dashed lines: Mubarak ousted (2011) and military coup (2013). Source: V-Dem v16.

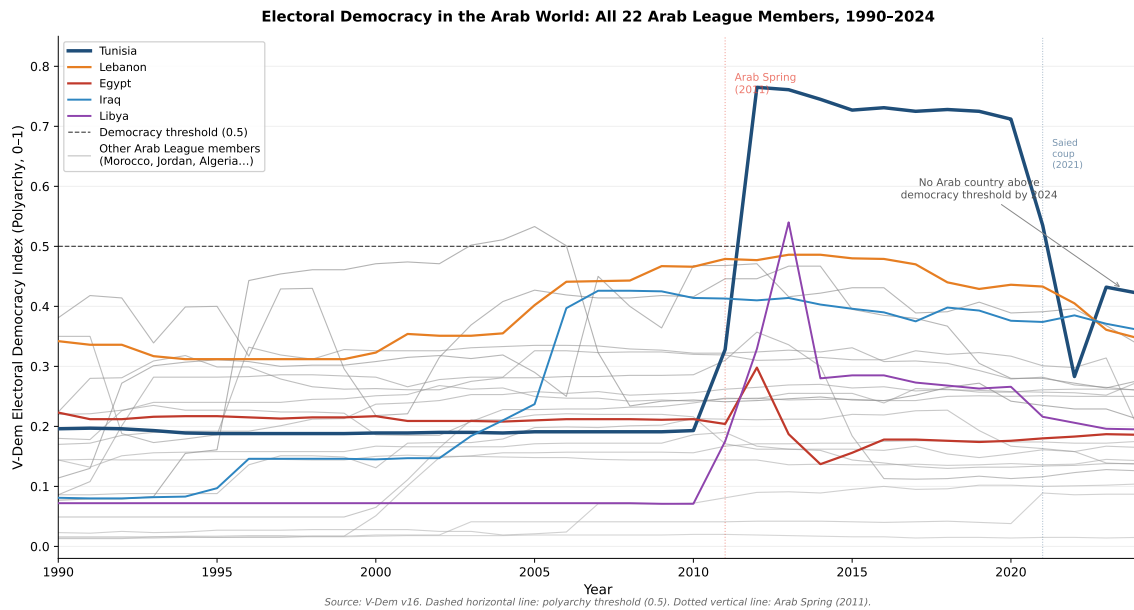


Figure 3: Electoral democracy (polyarchy) in Arab League member states, 1990–2024. Tunisia is above the descriptive threshold for most of 2012–2021; Egypt never crosses it. Vertical dashed line: Arab Spring (2011). Source: V-Dem v16.

is central to regime stability, particularly in rentier and patronage-based systems. Where the pre-transition order concentrated rents in a politically connected elite, the transition

is experienced by that elite as a loss measured against the previous status quo, not merely as a slower path to a common prosperity.

Third, reversals and consolidations coexist across economies with broadly comparable structural parameters. Countries with similar dividends and similar transition costs have followed divergent post-transition trajectories (Haggard and Kaufman, 2021). This suggests that regime outcomes are not pinned down by aggregate growth differentials alone, and invites a mechanism in which a behavioral parameter—here the loss-aversion coefficient—selects among outcomes that the structural primitives leave open.

These observations are motivating rather than dispositive. The model below does not test them; it formalises a channel consistent with them and derives its policy implication.

3 Environment

Time is discrete and infinite, $t = 0, 1, 2, \dots$

Definition 1 (The decision date). Date $t = 0$ is the first period in which a democratic framework is in force while the incumbent elite’s capacity to overturn it is still intact. “In force” means that civilian authority governs under an adopted constitutional framework, whether interim or permanent; “intact” means that the elite retains the coercive, economic, or bureaucratic means to restore authoritarian rule.

Definition 1 fixes the date against which “early” and “late” are measured throughout, and Section 9 applies it to specific transitions. It is not the date of the initial uprising, nor the date on which a permanent constitution replaces an interim one, unless those coincide with the start of the window just described.

Two windows must be kept apart, because they have different lengths and different content.

Definition 2 (Two windows). The *restoration-capacity window* is the interval over which the elite retains the means to overturn the democratic framework. The *behavioral vulnerability window* is the interval over which it would wish to use them, given the mechanism of this paper. The first is an institutional fact; the second is derived below, and Lemma 1 shows that under the baseline continuation path it coincides with the transition trough.

Capacity can therefore outlast vulnerability by many years, which is what makes the distinction useful: a concession that arrives inside the capacity window may nonetheless arrive after the mechanism has ceased to operate.

At $t = 0$ a single representative elite chooses between accommodating the transition (D) and resisting it, thereby restoring authoritarian rule (N). The origin of the transition is not modelled; the analysis is conditional on its initiation. There is no type heterogeneity, no private information, and no strategic counterparty in the baseline: the elite compares

two deterministic consumption streams and selects the one it prefers. The choice is treated as once-and-for-all; Lemma 1 below shows that this restriction is not binding in the present environment, because resistance is attractive only while the elite is in the loss region.

Restored autocracy is treated as absorbing in the baseline. Section 6 relaxes that assumption and shows that the absorbing case is the configuration most favourable to resistance.

The elite has effective discount factor

$$\tilde{\beta} = \beta(1 - \delta) \in (0, 1),$$

where β is the pure time-preference discount factor and δ is the per-period probability of exit from political relevance. The decomposition gives interpretive content—separating economic discounting from political risk—but the results depend only on $\tilde{\beta}$, which may equivalently be read as a primitive effective horizon.

The elite enters the decision with pre-transition consumption c_{-1} , inherited from the prior authoritarian regime. Under the two regimes, elite consumption follows

$$c_t^D = \begin{cases} c_{-1}(1 - \kappa), & t = 0, \\ c_{-1}(1 + G)^t, & t \geq 1, \end{cases} \quad (1)$$

$$c_t^N = c_{-1}(1 + a)^t, \quad t \geq 0, \quad (2)$$

with $G > a \geq 0$ and $\kappa \in (0, 1)$. The parameter κ is the elite-rent transition cost, measured as the loss the elite *anticipates at the decision date* rather than the loss realised later; G is the democratic dividend; a is the authoritarian counterfactual growth rate.

Two points on interpretation. First, $G > a$ is a comparison of growth rates and does *not* by itself make accommodation superior in present value: the present-value comparison is $B - \mathcal{L}_\kappa \geq 0$, equivalently $\Psi \geq 1$, where B and Ψ are defined in Section 5. The behavioral question posed in this paper arises only where the present-value comparison favours accommodation, and Corollary 1 states that restriction formally. Second, G is the growth rate of elite-relevant income, not of the rents the transition dissolves. The elite loses regulatory and patronage rents at the transition and thereafter holds diversified claims—firms, land, human capital—on an economy growing at G . Nothing below depends on the exact shape of the post-transition path: the elite’s problem depends on it only through the present-value statistic B , so any path that remains weakly above the reference from $t = 1$ onward and has the same B delivers the same results. That equivalence does not extend to Section 6, where the timing of losses under restoration enters separately.

Remark 1 (Interpretation of the one-period trough). The single-period dip in (1) should

be read as a reduced-form representation of the politically salient transition trough, not as a literal claim that all elite rents are restored after one period. What the model requires is that the transition entails an initial interval evaluated as a loss relative to the status quo, followed by a path that eventually exceeds it; the one-period normalisation isolates that structure in the simplest tractable form. A trough lasting several periods replaces $\mathcal{L}_\kappa$ by the discounted aggregate loss over the trough and leaves the threshold structure intact, though the aggregate depends on the profile of losses and not only on their total.

The elite has reference-dependent preferences over consumption (Kőszegi and Rabin, 2006). Period utility is the sum of an intrinsic logarithmic term and a gain-loss term,

$$U(c \mid r) = \log c + \eta n(c \mid r), \quad (3)$$

where r is the reference point, $\eta > 0$ is the gain-loss intensity, and

$$n(c \mid r) = \begin{cases} \log c - \log r, & c \geq r, \\ \lambda(\log c - \log r), & c < r, \end{cases} \quad (4)$$

with $\lambda \geq 1$ the loss-aversion coefficient. The gain-loss component is λ times steeper in the loss region than in the gain region. Lifetime utility under regime $\rho \in \{\mathbf{D}, \mathbf{N}\}$ is $V(\rho) = \sum_{t \geq 0} \tilde{\beta}^t U(c_t^\rho \mid r_t)$.

The once-and-for-all treatment of the elite's choice can be relaxed at no cost.

Lemma 1 (A Recurrent Veto Does Not Extend the Window). *Suppose the elite may restore authoritarian rule at any date $\tau \geq 0$, that restoration is costless, and that restoration at τ returns its consumption to the authoritarian counterfactual path $c_t^{\mathbf{N}} = c_{-1}(1+a)^t$ for all $t \geq \tau$. Let the trough last $T_0 \geq 1$ periods.*

- (a) *Resistance at any date $\tau \geq T_0$, that is after the trough, is strictly dominated by permanent accommodation.*
- (b) *With the one-period trough of (1), $T_0 = 1$, so the only candidate date is $\tau = 0$ and Theorem 1 characterises the choice.*
- (c) *With a trough of $T_0 > 1$ periods, as in Remark 1, the candidate dates are $\tau \in \{0, \dots, T_0 - 1\}$ and the elite resists if and only if it prefers resistance at the best of them.*

Proof. See Appendix A. □

The reason is that the trough is the only interval in which resistance is evaluated in the loss region. Once the elite is above its reference under either regime, the loss weight λ does not apply and the democratic path dominates the authoritarian one in level and

in growth, whatever the elite's loss aversion. Under the baseline continuation path the behavioral vulnerability window of Definition 2 therefore coincides with the trough, and the recurrent veto neither shortens nor extends it.

The force of that conclusion comes from the continuation path assumed: restoration returns the elite to the counterfactual it would have had without the transition, and recovers nothing more. Constitutional veto power, however, matters in practice precisely because it can recover rents. The following remark states what changes.

Remark 2 (Recoverable rents widen the window, but not behaviorally). Suppose restoration at τ delivers $c_t = c_{-1}(1+a)^t(1+\rho)$ for $t \geq \tau$, where $\rho \geq 0$ is the recoverable rent premium, and write $R = \log(1+\rho)$. Then resistance at $\tau \geq 1$ is preferred to permanent accommodation if and only if

$$\tau < \bar{\tau} \equiv \max \left\{ 0, \frac{R}{\log(1+G) - \log(1+a)} - \frac{\tilde{\beta}}{1 - \tilde{\beta}} \right\}. \quad (5)$$

The window collapses to the trough when $R = 0$, which is Lemma 1, and widens with the recoverable premium and with the horizon: at the configuration of Section 8, $\bar{\tau} = 0$ for $\rho \leq 0.1$, $\bar{\tau} \approx 2.3$ for $\rho = 0.2$, and $\bar{\tau} \approx 10$ for $\rho = 0.5$. The comparison in (5) involves no loss weight, because both paths lie above the benchmark from $t = 1$: late resistance driven by recoverable rents is a present-value phenomenon, not a behavioral one. The mechanism of this paper therefore applies to the trough, and a late intervention is evidence of recoverable rents rather than of loss aversion.

A breakdown occurring well after the transition cannot be explained by this mechanism as it stands. It requires either that the reference has adapted upward to the democratic path and a subsequent shock has opened a new loss region, or that the actors who overturn the regime are not the elite that accommodated it. Section 9 uses that distinction rather than eliding it.

4 The Status-Quo Reference Point

The conceptual heart of the model is the reference point. I take the reference to be the status-quo consumption level inherited from the previous regime,

$$r_t = c_{-1} \quad \text{for all } t \geq 0. \quad (6)$$

The interpretation is political-economic rather than purely consumption-based. The elite evaluates the regime decision against the rents, privileges, and status it held under the prior order. A democratic transition that erodes those rents is registered as a loss relative to the inherited benchmark; subsequent growth above the benchmark is registered as

a gain. Specification (6) encodes the idea that the politically decisive comparison is “relative to what we had,” which is the natural reference for an incumbent elite contemplating the surrender of entrenched advantage.

This specification differs from the rational-expectations reference of strict Kőszegi–Rabin personal equilibrium, under which the reference equals the anticipated consumption stream and a fully anticipated deterministic transition cost would be absorbed into the reference, eliminating the gain-loss penalty. Here the reference is anchored to the status quo and does not absorb the anticipated transition loss; the loss therefore retains its behavioral bite.

It should be stated plainly that (6) is a primitive of the model and not a consequence of loss aversion. The results below follow from loss aversion *together with* an entitlement reference anchored in the prior order, and the choice between that reference and an expectations-based one is a substantive disagreement about what an incumbent elite treats as its benchmark, not a modelling detail: under the expectations-based reference the mechanism does not operate at all. The interpretation offered here is that the benchmark is the position the previous constitutional order guaranteed the elite—rents, immunity, appointment rights—rather than a forecast of its own consumption, and that political entitlements of this kind are revised slowly. The paper does not derive that revision process from more primitive assumptions. What follows is accordingly an *entitlement-reference* model of elite resistance, and not a general model of loss-aversion-driven transition fragility: loss aversion alone, attached to a different benchmark, yields different results or none.

The benchmark is also collapsed into a single consumption index, whereas an elite may define its position in terms of legally protected assets, office and appointment rights, immunity from prosecution, relative standing, or expected future rents, and these need not move together. Two observations make the consumption index a defensible summary. First, the instruments a transition actually uses—amnesty, protection of titles, reserved posts—are convertible into consumption terms, which is what the model requires. Second, and more usefully, the composition of an elite’s demands is observable: an elite whose benchmark is the inherited entitlement asks for immunity, restitution, and the protection of positions it already held, whereas an elite whose benchmark is a forecast of its own future asks for a share of prospective growth. The former pattern is consistent with what has been documented of negotiated transitions and would support the reference used here; the latter would tell against it. The paper does not establish either pattern as a systematic comparative fact. What it does instead is to make the dependence of the results on the reference explicit and testable: Proposition 2 characterises the threshold for every adjustment speed χ , and Proposition 1 states a prediction that holds only if the reference is where the model places it. A reader who rejects the entitlement reference should read the paper as establishing what would follow if elites evaluated transitions

against it.

Remark 3 (Gradual reference adjustment). The fixed status-quo reference is the polar case $\chi = 0$ of the backward-looking rule

$$\log r_t = (1 - \chi) \log r_{t-1} + \chi \log c_t, \quad r_{-1} = c_{-1}, \quad \chi \in [0, 1), \quad (7)$$

in which the reference absorbs a fraction χ of current consumption in each period. Adjustment operates through two channels of opposite sign. It absorbs part of the transition trough into the reference, which shrinks the period-zero loss and favours accommodation; and it lets the reference track subsequent consumption growth, which erodes the gains that the faster-growing democratic path records relative to the authoritarian path, and so favours resistance. Neither channel dominates in general. The fixed reference is therefore a polar benchmark rather than an upper bound on the strength of the mechanism; Proposition 2 in Section 5 characterises the threshold under (7) in closed form.

Under (6), the elite is in the loss region exactly when consumption falls below c_{-1} . Define

$$\mathcal{L}_\kappa \equiv |\log(1 - \kappa)| = -\log(1 - \kappa) > 0, \quad (8)$$

the magnitude of the period-zero log-consumption shortfall under democracy. Under democracy the elite is in the loss region at $t = 0$, where $c_0^D = c_{-1}(1 - \kappa) < c_{-1}$, and in the gain region for $t \geq 1$, where $c_t^D = c_{-1}(1 + G)^t > c_{-1}$. Under authoritarian restoration the elite is exactly at the reference at $t = 0$ and in the gain region for $t \geq 1$.

5 The Resistance Threshold

Aggregating period utilities along each path and differencing yields the regime-choice criterion.

Lemma 2 (Value Differential). *Under the consumption paths (1)–(2), the status-quo reference (6), and preferences (3)–(4), the lifetime value differential between accommodation and resistance is*

$$V(D) - V(N) = \mathcal{L}_\kappa [(1 + \eta)\Psi - (1 + \eta\lambda)], \quad (9)$$

where

$$\Psi \equiv \frac{\tilde{\beta} [\log(1 + G) - \log(1 + a)]}{(1 - \tilde{\beta})^2 \mathcal{L}_\kappa}. \quad (10)$$

Proof. See Appendix A. □

The composite Ψ is the present value of the democratic growth dividend, expressed per unit of transition cost. It is the single sufficient statistic that compresses the technology and horizon parameters $(G, a, \tilde{\beta}, \kappa)$ relevant to the decision. The differential (9) is the

contest between the discounted dividend $(1 + \eta)\Psi$ and the loss-amplified transition cost $1 + \eta\lambda$, both measured in units of $\mathcal{L}_\kappa$.

Theorem 1 (Resistance Threshold). *Under the convention that the elite accommodates when indifferent, the representative elite accommodates the transition if and only if*

$$\lambda \leq \lambda^* \equiv \frac{(1 + \eta)\Psi - 1}{\eta}. \quad (11)$$

The threshold λ^* is admissible—that is, $\lambda^* \geq 1$, consistent with the loss-aversion restriction $\lambda \geq 1$ —if and only if $\Psi \geq 1$, i.e. the present value of the democratic dividend at least covers the transition cost.

Proof. The elite accommodates when (9) is non-negative, the weak inequality reflecting the tie-breaking convention. Since $\mathcal{L}_\kappa > 0$, this requires $(1 + \eta)\Psi \geq 1 + \eta\lambda$, equivalently $\lambda \leq [(1 + \eta)\Psi - 1]/\eta$. Admissibility $\lambda^* \geq 1$ is equivalent to $(1 + \eta)\Psi - 1 \geq \eta$, i.e. $\Psi \geq 1$. $\square$

The threshold admits the decomposition

$$\lambda^* = \Psi + \frac{\Psi - 1}{\eta}, \quad (12)$$

which separates a structural component Ψ from a behavioral correction $(\Psi - 1)/\eta$ that shrinks as the gain-loss intensity η rises.

Remark 4 (Effective loss aversion and identification). The marginal utility of log consumption is $1 + \eta$ just above the reference and $1 + \eta\lambda$ just below it. Define effective loss aversion as the ratio

$$\Lambda \equiv \frac{1 + \eta\lambda}{1 + \eta}.$$

By (9), $V(\mathbf{D}) - V(\mathbf{N}) = (1 + \eta)\mathcal{L}_\kappa(\Psi - \Lambda)$, so Theorem 1 is equivalently the condition $\Lambda \leq \Psi$: the elite accommodates if and only if its effective loss aversion does not exceed the present value of the dividend per unit of transition cost. The two parameters have distinct interpretations— η is the weight of gain-loss utility relative to intrinsic utility, and λ the asymmetry of gain-loss utility itself—but the accommodation decision depends on them only through Λ , in the baseline and under fragile restoration alike (Section 6). Regime choices therefore identify Λ , not (η, λ) separately. The two are separately identified only when the reference adjusts (Proposition 2), where η also governs how quickly gains are eroded as the reference tracks consumption, or from independent evidence on the two components of utility. The result $\partial\lambda^*/\partial\eta < 0$ in Remark 5 is the counterpart of this observation: holding Ψ fixed, a larger η requires a smaller λ to keep Λ at Ψ . Loss-aversion estimates obtained as ratios of the marginal valuation of losses to that of gains, such as the value 2.25 of [Tversky and Kahneman \(1992\)](#), correspond to Λ in this model

rather than to λ ; the two coincide only in the limit $\eta \rightarrow \infty$, in which intrinsic utility is negligible.

The admissibility condition $\Psi \geq 1$ is also the dividing line between resistance that any standard model predicts and resistance that only loss aversion can produce.

Corollary 1 (Present-Value and Behavioral Resistance). *(a) If $\Psi < 1$, the elite resists for every $\lambda \geq 1$ and $\eta \geq 0$, and resistance coincides with the choice of an elite without gain-loss utility ($\eta = 0$), which compares present values alone.*

(b) If $\Psi \geq 1$, an elite without gain-loss utility ($\eta = 0$) or without loss aversion ($\lambda = 1$) accommodates, and the elite resists if and only if $\lambda > \lambda^* \geq 1$.

Proof. With $\eta = 0$, (9) reduces to $\mathcal{L}_\kappa(\Psi - 1)$; with $\lambda = 1$ it reduces to $(1 + \eta)\mathcal{L}_\kappa(\Psi - 1)$. If $\Psi < 1$, then $(1 + \eta)\Psi - (1 + \eta\lambda) < \eta(1 - \lambda) \leq 0$ for all $\lambda \geq 1$. If $\Psi \geq 1$, both reduced differentials are non-negative, and the characterisation follows from Theorem 1. $\square$

Resistance in case (b) is *behavioral*: the elite rejects a transition that is not against its present-value interest (and is strictly in it when $\Psi > 1$), and would accept it if it weighted losses and gains symmetrically. All subsequent claims about the distinctive content of the mechanism refer to this region.

Remark 5 (Comparative Statics). In the admissible region $\Psi > 1$, the resistance threshold satisfies

$$\frac{\partial \lambda^*}{\partial \kappa} < 0, \quad \frac{\partial \lambda^*}{\partial G} > 0, \quad \frac{\partial \lambda^*}{\partial a} < 0, \quad \frac{\partial \lambda^*}{\partial \tilde{\beta}} > 0, \quad \frac{\partial \lambda^*}{\partial \delta} < 0, \quad \frac{\partial \lambda^*}{\partial \eta} < 0.$$

The proof is in Appendix A.

The signs are economically transparent. A larger transition cost κ deepens the over-weighted trough and lowers the threshold, making resistance more likely. A larger democratic dividend G raises the discounted gain and the threshold. A higher authoritarian counterfactual a improves the outside option and lowers the threshold. A longer effective horizon $\tilde{\beta}$ makes the elite more willing to wait for the dividend; equivalently, a higher political-exit hazard δ shortens the horizon and lowers the threshold. Finally, a higher gain-loss intensity η lowers λ^* in the admissible region, because by (12) a larger η erodes the behavioral cushion $(\Psi - 1)/\eta$.

The comparative statics in the structural parameters $(\kappa, G, a, \tilde{\beta})$ are shared, in sign, with a present-value model. What the mechanism adds is a discontinuity at the reference. Write

$$B \equiv \frac{\tilde{\beta} [\log(1 + G) - \log(1 + a)]}{(1 - \tilde{\beta})^2} = \Psi \mathcal{L}_\kappa$$

for the discounted dividend, so that the present-value gain from accommodation is $B - \mathcal{L}_\kappa$ and (9) reads $V(D) - V(N) = (1 + \eta)B - (1 + \eta\lambda)\mathcal{L}_\kappa$. Both expressions are linear in log-consumption shifts, which makes the value of a concession to the elite easy to state.

Proposition 1 (Reference Crossing). *Consider a proportional period-zero concession that raises the elite’s log consumption at $t = 0$ by $\varepsilon > 0$, with ε small enough that the elite does not cross its reference.*

- (a) *The concession raises $V(\mathbf{D}) - V(\mathbf{N})$ by $(1 + \eta\lambda)\varepsilon$ if the elite is below its reference at $t = 0$ and by $(1 + \eta)\varepsilon$ if it is weakly above. The marginal value of concessions is therefore discontinuous in the elite’s position at the reference, with jump ratio $\Lambda = (1 + \eta\lambda)/(1 + \eta)$.*
- (b) *Hence two groups receiving the same proportional concession, in the same form, at the same date, and with the same credibility, but of which one anticipates a loss $\kappa_1 > 0$ and the other is weakly above its benchmark, respond in the ratio Λ , and the ratio jumps at $\kappa = 0$ rather than varying smoothly with κ .*
- (c) *Suppose instead that preferences are reference-independent, with a smooth increasing felicity U and date weights that vary smoothly with κ —the class that covers accounts in which concessions matter because of discounting, limited commitment, political survival risk, or a short horizon, with no constraint switching at $\kappa = 0$. Then the marginal value of a proportional concession varies continuously with κ . With logarithmic utility it does not depend on κ at all, so the ratio in (b) equals one; with constant relative risk aversion γ it equals $(1 - \kappa_1)^{1-\gamma}$, continuous in κ_1 and equal to one at $\kappa_1 = 0$.*

Proof. See Appendix A. □

Proposition 1 is the paper’s discriminating prediction, and it is stated in this form deliberately. A prediction that early concessions are more effective than late ones does *not* discriminate: limited commitment, illiquidity, survival risk, and short horizons all imply it, and Section 7.2 adds a commitment reason of its own. Nor does the level of the ratio in part (b) discriminate on its own: curvature of reference-independent utility can generate a ratio above one, and a borrowing constraint that binds harder for those who lost more works in the same direction. What the smooth class cannot produce is a *discontinuity*: within it the effect of a concession varies continuously with the recipient’s anticipated loss, while under reference dependence it jumps as the recipient crosses the benchmark. The kink is in the elite’s utility, and what data would reveal is a jump in the marginal effect of a concession, not a kink in levels.

The scope of that claim should be stated exactly, because it is easy to overstate. A discontinuity rejects smooth reference-independent valuation. It does not identify reference dependence against all alternatives, because political environments generate non-smooth behaviour for reasons that have nothing to do with loss aversion: eligibility rules, thresholds at which credit or collateral constraints begin to bind, confiscation and prosecution

thresholds, minimum-winning-coalition and veto-point constraints, and office-holding restrictions. Each of these can produce a jump in the marginal response to a concession. What distinguishes the present mechanism is the *location* of the jump: at the recipient's own inherited position, not at an institutional threshold. The test therefore requires a setting in which the entitlement benchmark does not coincide with any legal or eligibility cut-off, and it should be reported as rejecting the smooth class and being consistent with reference dependence, rather than as identifying it. It holds form, date, and credibility fixed while varying position relative to the benchmark, and it requires observing which groups were placed below it: confiscation, exclusion from office, and the withdrawal of licences are recorded events.

The timing result follows as a corollary and is retained because it organises the discussion of policy, not because it discriminates.

Corollary 2 (Timing and Instrument Ranking at Constant Present Value). *(a) If the present-value gain $B - \mathcal{L}_\kappa$ is held fixed while the trough deepens ($dB = d\mathcal{L}_\kappa > 0$), the value of accommodation changes by $-\eta(\lambda - 1)d\mathcal{L}_\kappa \leq 0$: resistance becomes weakly more likely, strictly so if $\lambda > 1$.*

(b) An intervention that raises $B - \mathcal{L}_\kappa$ by $\varepsilon > 0$ raises $V(D) - V(N)$ by $(1 + \eta\lambda)\varepsilon$ if delivered as relief of the trough and by $(1 + \eta)\varepsilon$ if delivered as a larger long-run dividend.

The two effects in (b) coincide when $\lambda = 1$ or $\eta = 0$, and the effect in (a) then vanishes.

Proof. Differentiate $V(D) - V(N) = (1 + \eta)B - (1 + \eta\lambda)\mathcal{L}_\kappa$. □

5.1 Gradual Reference Adjustment

Under the adaptive rule (7), the elite remains in the loss region only at $t = 0$ under accommodation and is never in the loss region under resistance (Appendix A), so the threshold retains a closed form. Define the growth-tracking weight

$$k(\chi) \equiv \frac{\tilde{\beta}\chi}{1 - \tilde{\beta} + \tilde{\beta}\chi} \in [0, 1).$$

Proposition 2 (Resistance Threshold under Gradual Reference Adjustment). *Under (7) with $\chi \in [0, 1)$:*

(a) The elite accommodates if and only if $\lambda \leq \lambda^*(\chi)$, where

$$\lambda^*(\chi) = \Psi - k(\chi)(\Psi - 1) + \frac{\Psi - 1}{\eta(1 - \chi)}, \quad (13)$$

which reduces to (11) at $\chi = 0$.

(b) Suppose $\Psi > 1$. Then $\lambda^*(\chi)$ is strictly convex on $[0, 1)$ and diverges as $\chi \rightarrow 1$.

(c) Suppose $\Psi > 1$. If $\eta \leq (1 - \tilde{\beta})/\tilde{\beta}$, $\lambda^*(\chi)$ is strictly increasing. If $\eta > (1 - \tilde{\beta})/\tilde{\beta}$, $\lambda^*(\chi)$ is U-shaped with unique minimiser

$$\chi_{\min} = \frac{\sqrt{\eta\tilde{\beta}(1 - \tilde{\beta})} - (1 - \tilde{\beta})}{\sqrt{\eta\tilde{\beta}(1 - \tilde{\beta})} + \tilde{\beta}} \in (0, 1), \quad (14)$$

and $\lambda^*(\chi) < \lambda^*(0)$ for every $\chi \in (0, \hat{\chi})$, where

$$\hat{\chi} = \frac{\eta\tilde{\beta} - (1 - \tilde{\beta})}{\tilde{\beta}(1 + \eta)} \in (0, 1). \quad (15)$$

(d) Suppose $\Psi > 1$. For every λ , the set of adjustment speeds at which the elite resists is an interval. If $\lambda > \lambda^*(0)$, it is $[0, \bar{\chi}(\lambda))$, where $\bar{\chi}(\lambda) \in (0, 1)$ is strictly increasing in λ and, when $\eta > (1 - \tilde{\beta})/\tilde{\beta}$, exceeds $\hat{\chi}$.

Proof. See Appendix A. □

The two terms in which χ enters (13) are the two channels of Remark 3: $-k(\chi)(\Psi - 1)$ is the growth-tracking channel, which lowers the threshold, and $(\Psi - 1)/[\eta(1 - \chi)]$ is the trough-absorption channel, which raises it. Three features of Proposition 2 bear on the robustness of the mechanism. First, the fixed reference is not the case most favourable to resistance: whenever $\eta > (1 - \tilde{\beta})/\tilde{\beta}$ —satisfied for any $\eta \geq 1$ when $\tilde{\beta} > 1/2$ —moderate adjustment strengthens the mechanism, and the threshold returns to its fixed-reference value only at $\hat{\chi}$. Second, $\chi_{\min}$ and $\hat{\chi}$ depend only on $(\eta, \tilde{\beta})$ and not on Ψ : the non-monotonicity is a property of preferences and horizon, not of technology. Third, part (d) answers the question of how slowly the reference must adjust for the mechanism to operate: an elite that resists under the fixed reference continues to resist for every adjustment speed below $\bar{\chi}(\lambda)$, and $\bar{\chi}(\lambda) > \hat{\chi}$. At the benchmark of Section 8, $\chi_{\min} = 1/6$, $\lambda^*(\chi_{\min}) \approx 2.13$, and $\hat{\chi} = 0.375$; an elite with $\lambda = 2.25$, which lies marginally inside the accommodation region under the fixed reference, resists for all adjustment speeds between approximately 0.0005 and 0.374. For any finite λ , resistance is overturned at $\bar{\chi}(\lambda) < 1$, and $\bar{\chi}(\lambda) \rightarrow 1$ only as $\lambda \rightarrow \infty$; at the benchmark, $\bar{\chi}(2.5) \approx 0.52$, $\bar{\chi}(3) \approx 0.66$, and $\bar{\chi}(5) \approx 0.84$. In the limit $\chi \rightarrow 1$ gain-loss utility vanishes and the choice reduces to the present-value comparison of Corollary 1.

6 Fragile Authoritarian Restoration

The baseline treats restored autocracy as absorbing. Restorations are, however, themselves exposed to renewed mobilisation, splits within the coercive apparatus, and external

intervention. This section relaxes the assumption and asks how the fragility of restoration changes the threshold. The answer is not only that fragility favours accommodation, as a present-value model would also predict, but that it interacts with loss aversion in a way that can reverse the sign of the behavioral effect.

6.1 Environment with a Restoration Hazard

Suppose that, after resistance at $t = 0$, the restored autocracy is displaced in each period $t \geq 1$ with hazard $\pi \in [0, 1]$, conditional on having survived to $t - 1$; displacement is absorbing. Upon displacement the elite loses its rents: its consumption falls to $c_{-1}(1 - \xi)$, $\xi \in (0, 1)$, in the period of displacement, and the shortfall then decays geometrically at rate $\psi \in (0, 1]$, so that k periods after displacement log consumption falls short of the reference by $\psi^k \Xi$ with $\Xi \equiv -\log(1 - \xi)$. The case $\psi = 1$ is a permanent loss and $\psi < 1$ a loss from which the elite partially recovers; the results below are stated for general ψ precisely because the permanent case is the most favourable to the reversal in part (b) of Proposition 3. Accommodated democracy remains absorbing, so that the analysis isolates the asymmetry to which the baseline is exposed. The reference remains the status quo c_{-1} (the case $\chi = 0$), so that a displaced elite is in the loss region. Write $\tilde{\beta}_\pi \equiv \tilde{\beta}(1 - \pi)$ and

$$B_\pi \equiv \frac{\tilde{\beta} \log(1 + G)}{(1 - \tilde{\beta})^2} - \frac{\tilde{\beta}_\pi \log(1 + a)}{(1 - \tilde{\beta}_\pi)^2}, \quad E_\pi \equiv \frac{\tilde{\beta} \pi \Xi}{(1 - \tilde{\beta}\psi)(1 - \tilde{\beta} + \tilde{\beta}_\pi)}. \quad (16)$$

B_π is the discounted dividend of accommodation relative to a restoration whose gains accrue only while it survives, and E_π is the discounted expected displacement loss under restoration. Both reduce to the baseline at $\pi = 0$: $B_0 = B$ and $E_0 = 0$.

Lemma 3 (Value Differential under Fragile Restoration). *The lifetime value differential between accommodation and resistance is*

$$V(\mathbf{D}) - V(\mathbf{N}) = (1 + \eta)B_\pi - (1 + \eta\lambda)M_\pi, \quad M_\pi \equiv \mathcal{L}_\kappa - E_\pi. \quad (17)$$

Proof. See Appendix A. □

The loss-aversion coefficient now multiplies the *net* loss exposure M_π : the transition trough, which accommodation imposes, less the expected displacement loss, which resistance risks. Define the critical hazard

$$\pi^c \equiv \frac{(1 - \tilde{\beta}\psi)(1 - \tilde{\beta})\mathcal{L}_\kappa}{\tilde{\beta}[\Xi - (1 - \tilde{\beta}\psi)\mathcal{L}_\kappa]} \quad \text{when } \tilde{\beta}\Xi > (1 - \tilde{\beta}\psi)\mathcal{L}_\kappa, \quad (18)$$

and $\pi^c \equiv 1$ otherwise; $M_\pi > 0$ if and only if $\pi < \pi^c$.

Proposition 3 (Resistance Threshold under Fragile Restoration). *(a) If $\pi < \pi^c$, the elite accommodates if and only if*

$$\lambda \leq \lambda^{**} \equiv \frac{(1 + \eta)\Psi_\pi - 1}{\eta}, \quad \Psi_\pi \equiv \frac{B_\pi}{M_\pi}. \quad (19)$$

The threshold satisfies $\lambda^{**} \geq \lambda^*$, with equality if and only if $\pi = 0$, and, for $\pi > 0$, is strictly increasing in π , in ξ , and in the persistence ψ of the displacement loss.

(b) If $\pi \geq \pi^c$, the elite accommodates for every $\lambda \geq 1$, and $V(\mathbf{D}) - V(\mathbf{N})$ is non-decreasing in λ : loss aversion strengthens, rather than weakens, the elite's preference for accommodation.

Proof. See Appendix A. □

Part (a) answers the question of direction and magnitude that the absorbing benchmark raises. Fragility of restoration raises the threshold through two channels: it shortens the expected duration of the authoritarian gains ($B_\pi > B$), and it attaches a loss to resistance ($E_\pi > 0$), which loss aversion amplifies. The magnitude is large. At the benchmark of Section 8 with $\xi = 0.3$ and a permanent loss ($\psi = 1$), a displacement hazard of one per cent per period raises the threshold from 2.25 to 3.26, and a hazard of two per cent raises it to 4.73 (Figure 6b). Persistence matters: at $\pi = 0.02$ the threshold falls from 4.73 to 3.34 when the loss decays at $\psi = 0.7$. The absorbing specification is therefore the configuration most favourable to resistance, and the baseline threshold should be read as a lower bound.

Part (b) is the substantive new content, and it is a conditional possibility rather than a general prediction: the reversal requires the joint configuration $E_\pi \geq \mathcal{L}_\kappa$, that is a hazard, a displacement loss, and a persistence of that loss large enough relative to the trough and the horizon. Where that configuration holds, loss aversion cuts both ways: the same asymmetry that overweights the transition trough also overweights the loss that a failed restoration would inflict. Which effect dominates depends on whether the trough exceeds the discounted expected displacement loss. At the benchmark with $\xi = 0.3$, $\pi^c = 0.0625$ under a permanent loss and 0.196 when the loss decays at $\psi = 0.7$: the reversal requires both a fragile restoration and a displacement loss that persists, and it should be read as conditional on that structure rather than as a general prediction. The loss-aversion mechanism therefore predicts that behavioral resistance is concentrated where restored autocracy is secure—where the coercive apparatus has both the capacity and the will to hold power—and that, where restoration is fragile, loss aversion strengthens rather than weakens the elite's preference for accommodation. A present-value model predicts neither the concentration nor the sign reversal, because it has no λ to interact with π .

6.2 Endogenous Fragility

The hazard π is not a free parameter. Two primitives plausibly govern it. Citizens who have experienced part of the democratic dividend and see it withdrawn have more to mobilise for, and the coercive capacity of the restored regime deters and represses mobilisation (Bellin, 2012). A convenient specification is

$$\pi(G, \varphi) = 1 - \exp\left[-\frac{\theta_G (G - a)}{\varphi}\right], \quad \theta_G > 0, \varphi > 0, \quad (20)$$

where φ is the repression capacity of the restored regime. Any specification with $\pi_G > 0$ and $\pi_\varphi < 0$ delivers the following result.

Corollary 3 (Primitives of Fragility). *Under (20) and $\pi < \pi^c$, the threshold λ^{**} is strictly increasing in the democratic dividend G and strictly decreasing in repression capacity φ .*

Proof. By Proposition 3(a), λ^{**} is strictly increasing in π ; under (20), π is strictly increasing in G and strictly decreasing in φ . In addition, G raises B_π directly, since $\partial B_\pi / \partial G = \tilde{\beta} / [(1 + G)(1 - \tilde{\beta})^2] > 0$, which reinforces the effect. $\square$

The dividend therefore acts twice: it raises the value of accommodation and the fragility of any restoration that would forgo it. Two channels are deliberately excluded. The hazard is not allowed to depend on a settlement that the elite was offered and refused: a refused offer can affect citizens' behaviour only through what it reveals, and modelling that requires the inference structure this paper sets aside. Nor is democracy, once accommodated, exposed to a hazard of its own; that asymmetry is the price of isolating the elite's decision, and Lemma 1 states what it implies about late breakdowns. Whether an offer large enough to secure accommodation is *feasible* is the subject of the next section.

7 Transitional Compensation and Constitutional Protections

Democracies facing an elite that can still overturn them do not rely on a single instrument. They pay cash, they legislate amnesties and reconciliation procedures, and they entrench protections in the constitution itself. This section treats the three as alternative ways of conceding present value to the elite, and asks which of them a constitutional designer acting for citizens should use. The answer turns on two things: how much of the concession's value falls in the elite's loss region, and what each instrument costs citizens.

7.1 Three Instruments

Let the constitutional settlement consist of

- (i) a *cash transfer* $s \geq 0$ paid at $t = 0$, financed by taxation and borrowing;
- (ii) a *transitional protection* $\sigma \geq 0$: a one-off legal guarantee—amnesty, immunity, protection of titles acquired under the prior order, a reconciliation procedure—whose value to the elite is concentrated in the transition period;
- (iii) a *recurring constitutional privilege* $q \in [0, \bar{q}]$: an entrenched protection of elite positions—reserved powers, supermajority requirements, protected sectors—that raises elite consumption in every period by the factor $1 + q$.

What distinguishes (ii) from (iii) is not legal duration but where the value falls. An amnesty is legally permanent: prosecution is foreclosed for good. What is temporary is the risk it removes, since exposure to prosecution, confiscation, and the revision of titles is concentrated in the aftermath of a transition, so the elite’s valuation of the guarantee is concentrated there too. That is why an amnesty maps to σ and a reserved power maps to q , and why “transitional” rather than “expiring” is the accurate label.

Write $m \equiv \log(1 + q)$ and let the entrenchment distort policy, so that the dividend becomes $G(q)$ with $G'(q) < 0$ and $B(q) \equiv \tilde{\beta}[\log(1 + G(q)) - \log(1 + a)]/(1 - \tilde{\beta})^2$ strictly decreasing. Throughout, entrenchment is assumed not to eliminate the dividend: $G(q) > a$, and hence $B(q) > 0$, for every $q \in [0, \bar{q}]$. Elite consumption is then

$$c_0^D = c_{-1}(1 - \kappa + s + \sigma)(1 + q), \quad c_t^D = c_{-1}(1 + G(q))^t(1 + q), \quad t \geq 1,$$

and the effective trough is $\mathcal{L}_\kappa(s, \sigma) \equiv -\log(1 - \kappa + s + \sigma)$, with $s + \sigma < \kappa$ so that the elite remains in the loss region at $t = 0$. For later reference, write $x \equiv s + \sigma$ for total period-zero relief and, setting $q = 0$,

$$\Psi(x) = \frac{\tilde{\beta} [\log(1 + G) - \log(1 + a)]}{(1 - \tilde{\beta})^2 \mathcal{L}_\kappa(x)}, \quad \lambda^*(x) = \frac{(1 + \eta)\Psi(x) - 1}{\eta}. \quad (21)$$

Remark 6 (Relief of the trough raises the threshold). $\lambda^*(x)$ is strictly increasing in x on $[0, \kappa)$, since $\mathcal{L}_\kappa(x)$ is strictly decreasing in x and λ^* is strictly decreasing in $\mathcal{L}_\kappa$ (Appendix A).

Lemma 4 (Value Differential with Instruments). *The value differential is piecewise linear*

in the shifts the instruments deliver:

$$V(\mathbf{D}) - V(\mathbf{N}) = \begin{cases} -(1 + \eta\lambda) [\mathcal{L}_\kappa(s, \sigma) - m] + (1 + \eta) \left[B(q) + \frac{\tilde{\beta} m}{1 - \tilde{\beta}} \right], & m < \mathcal{L}_\kappa(s, \sigma), \\ (1 + \eta) \left[m - \mathcal{L}_\kappa(s, \sigma) + B(q) + \frac{\tilde{\beta} m}{1 - \tilde{\beta}} \right], & m \geq \mathcal{L}_\kappa(s, \sigma). \end{cases} \quad (22)$$

Proof. See Appendix A. □

The first branch applies while the settlement leaves the elite below its reference at $t = 0$; the second applies once the instruments lift it above. Within each branch the differential is linear in the shifts delivered, which is what makes the design problem tractable. The branch boundary is $m = \mathcal{L}_\kappa(s, \sigma)$ and not $s + \sigma = \kappa$, so it can be reached by entrenchment as well as by relief of the trough. The next result shows that the boundary never binds on the design problem.

Lemma 5 (Crossing the Reference Is Never the Binding Case). *If $m \geq \mathcal{L}_\kappa(s, \sigma)$ then $V(\mathbf{D}) - V(\mathbf{N}) > 0$ for every $\lambda \geq 1$ and $\eta \geq 0$. Consequently the accommodation constraint can bind only where $m < \mathcal{L}_\kappa(s, \sigma)$, and any settlement that satisfies it with equality lies strictly in that region, where the first branch of (22) applies.*

Proof. See Appendix A. □

Lemma 5 is the reason the analysis below can work with the first branch alone: a settlement generous enough to lift the elite above its inherited benchmark has already secured accommodation, whatever the elite's loss aversion, so a designer minimising the cost of accommodation never goes that far.

7.2 The Design Problem and the Credibility of Promises

A representative citizen has present-value welfare, measured relative to authoritarian restoration and in units of pre-transition output,

$$W(\{s_t\}, \sigma, q) = B^c(q) - \nu \left[s_0 + \sum_{t \geq 1} \tilde{\beta}^t s_t \right] - c_\sigma \sigma, \quad (23)$$

where $B^c(q)$ is the citizens' democratic surplus, strictly decreasing in entrenchment; $\nu \equiv c_{-1}/Y$ is the elite's pre-transition share of output, so that a transfer of s in any period costs citizens νs in that period, future payments included; and $c_\sigma > 0$ is the cost citizens attach to legal protection of the outgoing elite—forgone recovery of assets, impunity, and the legitimacy cost of both. Restoration yields citizens $W = 0$ by normalisation, and $B^c(0) > 0$, so citizens prefer democracy without protections to restoration.

The government acts for the citizen, maximising (23), which prices future transfers as well as current ones. The order of moves is: (1) the constitutional settlement (σ, q) is enacted and the period-zero transfer s_0 is paid or legally vested, and a schedule $\{s_t\}_{t \geq 1}$ may be announced; (2) the elite accommodates or resists; (3) after accommodation, in each period $t \geq 1$ the government chooses whether to pay s_t . Constitutional provisions are not re-chosen in stage (3): amending them requires the procedure that entrenched them.

Lemma 6 (Only Vested Concessions Are Credible). *In every subgame-perfect equilibrium of this game, $s_t = 0$ for all $t \geq 1$, and the elite values an announced schedule $\{s_t\}_{t \geq 1}$ at zero. Concessions that bind the government in later periods must therefore be vested at $t = 0$ or constitutionally entrenched.*

Proof. See Appendix A. □

Lemma 6 is the commitment problem of Acemoglu and Robinson (2001) applied to the transfer rather than to the policy, and it identifies the role of the constitution in this model: entrenchment is the technology that makes a promise for $t \geq 1$ credible, at the cost of distorting policy forever. It also removes the temptation to read the model as claiming that a democracy can always buy off an elite. It cannot: it can pay now, protect now, or entrench, and each of the three is limited.

7.3 Behavioral Leverage and the Optimal Instrument Mix

Measure each instrument by the present value it delivers to the elite in log-consumption units: P_s and P_σ for cash and transitional protection, each equal to the reduction in $\mathcal{L}_\kappa$, and $P_q = m/(1 - \tilde{\beta})$ for entrenchment. Define the *behavioral leverage* of an instrument as the relaxation of the accommodation constraint per unit of present value delivered.

Proposition 4 (Leverage of Concessions). *The behavioral leverage of the three instruments is*

$$\Lambda_s = \Lambda_\sigma = 1 + \eta\lambda, \quad \Lambda_q = (1 - \tilde{\beta})(1 + \eta\lambda) + \tilde{\beta}(1 + \eta) = (1 + \eta) + (1 - \tilde{\beta})\eta(\lambda - 1).$$

Hence $\Lambda_\sigma - \Lambda_q = \tilde{\beta}\eta(\lambda - 1) \geq 0$, with equality if and only if $\lambda = 1$, and the dominance margin

$$\frac{\Lambda_\sigma}{\Lambda_q} = \frac{1 + \eta\lambda}{(1 + \eta) + (1 - \tilde{\beta})\eta(\lambda - 1)}$$

is strictly increasing in λ and in $\tilde{\beta}$. In terms of effective loss aversion it takes the form

$$\frac{\Lambda_\sigma}{\Lambda_q} = \frac{\Lambda}{1 + (1 - \tilde{\beta})(\Lambda - 1)}, \quad \Lambda = \frac{1 + \eta\lambda}{1 + \eta},$$

which depends only on Λ and $\tilde{\beta}$ and not on η and λ separately. The quantities Λ_s , Λ_σ , and Λ_q are gross of the policy distortion. Because entrenchment also lowers the dividend, its net leverage is

$$\Lambda_q^{\text{eff}} = \Lambda_q + (1 + \eta)(1 - \tilde{\beta})(1 + q) B'(q) < \Lambda_q,$$

and it is Λ_q^{eff} , not Λ_q , that enters the exact optimality conditions of Proposition 5. All statements are made on the branch $m < \mathcal{L}_\kappa(s, \sigma)$, which by Lemma 5 is where the accommodation constraint binds; on the other branch all three instruments have leverage $1 + \eta$ and the ranking is vacuous.

Proof. See Appendix A. □

The economics is that a transitional protection concentrates its entire value in the single period in which the elite is below its reference, where a unit of consumption is worth $1 + \eta\lambda$, whereas entrenchment spreads its value over the whole path, of which only the fraction $1 - \tilde{\beta}$ falls in the loss region. At the illustrative configuration of Section 8 the margin is 1.44: per unit of present value conceded, a protection whose value falls in the trough relaxes the accommodation constraint 44 per cent more than a privilege spread over the whole path, and more than that once the dividend cost of entrenchment is counted. With $\eta = 2$, $\lambda = 3$, and $\tilde{\beta} = 0.85$ the margin is 1.94.

Let $D \equiv (1 + \eta\lambda)\mathcal{L}_\kappa - (1 + \eta)B(0) > 0$ denote the accommodation gap at $(s, \sigma, q) = (0, 0, 0)$, and let p_j denote the marginal citizen cost of instrument j per unit of present value delivered, so that the ranking of the two trough instruments is a pure price comparison: $p_s = \nu(1 - \kappa + s + \sigma)$ and $p_\sigma = c_\sigma(1 - \kappa + s + \sigma)$, whence $p_s/p_\sigma = \nu/c_\sigma$, and $p_q = -(1 - \tilde{\beta}) B'(q)(1 + q)$.

Two cases must be distinguished. If

$$-(1 + q)B'(q) < \frac{\Lambda_q}{(1 + \eta)(1 - \tilde{\beta})} \quad \text{for all } q \in [0, \bar{q}], \quad (24)$$

then $\Lambda_q^{\text{eff}} > 0$ and entrenchment is a concession: its direct value to the elite outweighs the dividend it destroys. If (24) fails at some q , then $\Lambda_q^{\text{eff}} \leq 0$ there, so entrenchment both costs citizens and tightens the accommodation constraint, and it is never used. The results below therefore hold in either case, with the ranking of Proposition 5 holding trivially where (24) fails. At the illustrative configuration of Section 8, (24) holds whenever the marginal dividend cost of entrenchment satisfies $|G'(q)| < 0.295$ at $q = 0$.

Proposition 5 (Optimal Constitutional Settlement). *Consider $\max_{s, \sigma, q} W(s, \sigma, q)$ subject to $V(D) \geq V(N)$, the liquidity bound $s \leq \varsigma/\nu$, $s + \sigma < \kappa$, $q \in [0, \bar{q}]$, and $m \leq \mathcal{L}_\kappa(s, \sigma)$; by Lemma 5 the last constraint does not bind at an optimum, so it is imposed without loss.*

- (a) If $D \leq 0$, that is $\lambda \leq \lambda^*$, the solution is $(s, \sigma, q) = (0, 0, 0)$: protections are pure cost and are used only to avert resistance.
- (b) If $D > 0$, the accommodation constraint binds, and the following first-order characterisation holds: any two instruments used at an interior optimum have the same marginal cost per unit of relaxation, $p_j/\Lambda_j^{\text{eff}}$; an instrument whose marginal ratio is strictly higher is not used, and one whose bound binds has a strictly lower ratio. Because $p_s/p_\sigma = \nu/c_\sigma$ does not vary with the amounts conceded and $\Lambda_s^{\text{eff}} = \Lambda_\sigma^{\text{eff}}$, the comparison of cash and transitional protection is global: the cheaper of the two is used up to its bound before the other is used at all. For entrenchment the characterisation is local unless $p_q/\Lambda_q^{\text{eff}}$ is non-decreasing in q , in which case it is global as well.
- (c) If citizens price a unit of present value conceded through transitional protection and through entrenchment equally, $p_\sigma = p_q$, the optimum uses transitional protection alone up to its bound. Entrenchment enters only if $p_q/p_\sigma \leq \Lambda_q^{\text{eff}}/\Lambda_\sigma$. Gross of the dividend distortion, that admissible price ratio falls as Λ or $\tilde{\beta}$ rises, so the price advantage entrenchment must offer grows with effective loss aversion and with the elite's horizon. The same conclusion holds net of the distortion under the sufficient condition
- $$-(1+q)B'(q) \leq \min \left\{ \frac{\tilde{\beta}}{1-\tilde{\beta}}, \Lambda - 1 \right\}, \quad (25)$$
- and it can fail without it, since the distortion term does not scale with λ .
- (d) Cash and transitional protection have the same leverage and differ only in price and in feasibility: both are limited by $s + \sigma < \kappa$, but only cash faces the fiscal bound $s \leq \varsigma/\nu$, which tightens as rent concentration ν rises.

Proof. See Appendix A. □

Proposition 5 is the paper's constitutional result, and its content should be stated precisely, because two comparisons are doing different work. The behavioral comparison is between concessions delivered in the trough and privileges spread over the whole path: a unit of present value conceded in the trough has strictly greater leverage, by the factor of Proposition 4, and that comparison is independent of prices. The comparison *within* the trough—cash against transitional protection—is not behavioral at all, since the two have identical leverage; it is settled entirely by the price ratio ν/c_σ , so cash is used first where citizens object less to paying than to granting impunity, and protection first where they object less to impunity. The ordering the model delivers is therefore: concede in the trough, through whichever trough instrument citizens price more cheaply, and turn to recurring privilege last. That ranking is global for the two trough instruments and local for entrenchment, unless $p_q/\Lambda_q^{\text{eff}}$ is monotone.

So stated, the result gives a formal reason for a pattern the comparative literature records descriptively—that transitional bargains rely on amnesties and transitional guarantees rather than on permanent constitutional privileges (O’Donnell and Schmitter, 1986; Linz and Stepan, 1996)—and it distinguishes those settlements from the elite-biased constitutions of Albertus and Menaldo (2018), which in this model are the expensive instrument, chosen when the cheap ones are exhausted or unavailable.

7.4 When No Settlement Exists

Nothing guarantees that the two sides can both be satisfied. Let $\mathcal{C}(D)$ be the minimum citizen cost of closing the accommodation gap D subject to the bounds of Proposition 5; $\mathcal{C}$ is non-decreasing, and $\mathcal{C}(D) = \infty$ where the bounds make D unreachable. No curvature of $\mathcal{C}$ is needed for what follows.

Proposition 6 (Constitutional Infeasibility). *A settlement acceptable to both sides exists if and only if $\mathcal{C}(D) \leq B^c(0)$, the citizens’ surplus from democracy without protections. Otherwise every settlement that secures accommodation leaves citizens worse off than authoritarian restoration, so no settlement in the admissible instrument set satisfies both participation constraints and the transition breaks down. Holding the remaining primitives fixed, the accommodation gap D increases in the elite’s loss aversion λ and in the transition cost κ ; the cash bound ς/ν tightens and the marginal price of cash rises as rent concentration ν rises; and the attainable set expands in transfer capacity ς and in the permissible entrenchment $\bar{q}$. The infeasible region therefore expands in λ , κ , and ν and contracts in ς , $\bar{q}$, and the citizens’ surplus.*

Proof. See Appendix A. □

Proposition 6 locates a bargaining frontier: within the admissible instrument set, concessions can go as far as the citizens’ surplus from democracy and no further. The statement is positive and conditional on the reduced-form citizen objective (23); the normative content sits entirely in the exogenous objects $B^c(\cdot)$, c_σ , $\bar{q}$ and ς , which the model does not derive from electoral institutions or political preferences. It is therefore not a constitutional norm about how far a democracy *should* go in protecting those it replaces. What it does say is that the cheapest instruments are the ones that offend most directly against accountability—impunity for the outgoing order—so the frontier binds on the dimension where it is least comfortable, and that where it binds the model predicts breakdown rather than settlement, which is a more disciplined prediction than the unbounded compensation result of the baseline.

Remark 7 (The uncompensated threshold as a special case). Setting $\sigma = q = 0$ and treating s as unconstrained recovers the earlier statement that any elite can be retained:

as $s \uparrow \kappa$, $\mathcal{L}_\kappa(s, 0) \rightarrow 0$ and the threshold diverges. That statement is a feasibility limit, not a policy conclusion, and Propositions 5 and 6 replace it.

8 Numerical Illustration

The configuration used below is an illustration, not a calibration. No parameter is estimated from any of the transitions discussed in Section 9, and the laboratory estimates of loss aversion available in the literature are measured on individual choices rather than on the decisions of political elites. The configuration is $\tilde{\beta} = 0.80$, $G = 0.05$, $a = 0.02$, $\kappa = 0.30$, which give $\mathcal{L}_\kappa = |\log 0.70| = 0.357$ and $\Psi = 1.625$. By Remark 4, the accommodation condition is $\Lambda \leq \Psi$, so the illustration places the frontier at effective loss aversion $\Lambda^* = 1.625$; with $\eta = 1$ this corresponds to $\lambda = 2.25$, which is why results are also reported in λ . Nothing below depends on identifying η and λ separately. Read as an estimate of Λ rather than of λ —which by Remark 4 is the correct mapping for a ratio of marginal valuations—the experimental value 2.25 would place the elite outside the accommodation region at this configuration, implying $\lambda = 3.5$ at $\eta = 1$ and relief of about a quarter of the trough to accommodate it. That is a further reason to read the configuration as locating the frontier rather than as classifying any actual elite.

8.1 The Threshold across Structural Configurations

Figure 4 plots level sets of $\lambda^*(G, \kappa)$ at $\eta = 1$; the shaded region is the present-value region $\Psi < 1$ of Corollary 1. The practical content is the slope of the level sets: along the benchmark level set, an increase of the transition cost from 0.30 to 0.35 must be offset by an increase of the dividend from 5.0 to about 5.6 per cent per period to keep the same elite on the accommodation side of the frontier.

8.2 Concessions: How Much, and in What Form

Table 1 reports the relief of the trough x^* required to accommodate an elite with effective loss aversion Λ , and Figure 5 plots the frontier. Near the frontier little is needed: relieving 6 per cent of the trough accommodates $\Lambda = 1.75$ and 16 per cent accommodates $\Lambda = 2$. The requirement then rises steeply, to half the trough at $\Lambda = 3.5$ and two-thirds at $\Lambda = 5.5$. As a share of output the cost of cash relief is νx^* : with an elite share of output $\nu = 0.2$, accommodating $\Lambda = 2$ costs about one per cent of pre-transition output, which is the order of magnitude at which the liquidity bound of Proposition 5 begins to bind.

The form of the concession matters as much as its size. Table 2 reports the dominance margin $\Lambda_\sigma/\Lambda_q = \Lambda/[1 + (1 - \tilde{\beta})(\Lambda - 1)]$ of Proposition 4: the factor by which a transitional protection relaxes the accommodation constraint more than permanent entrenchment, per

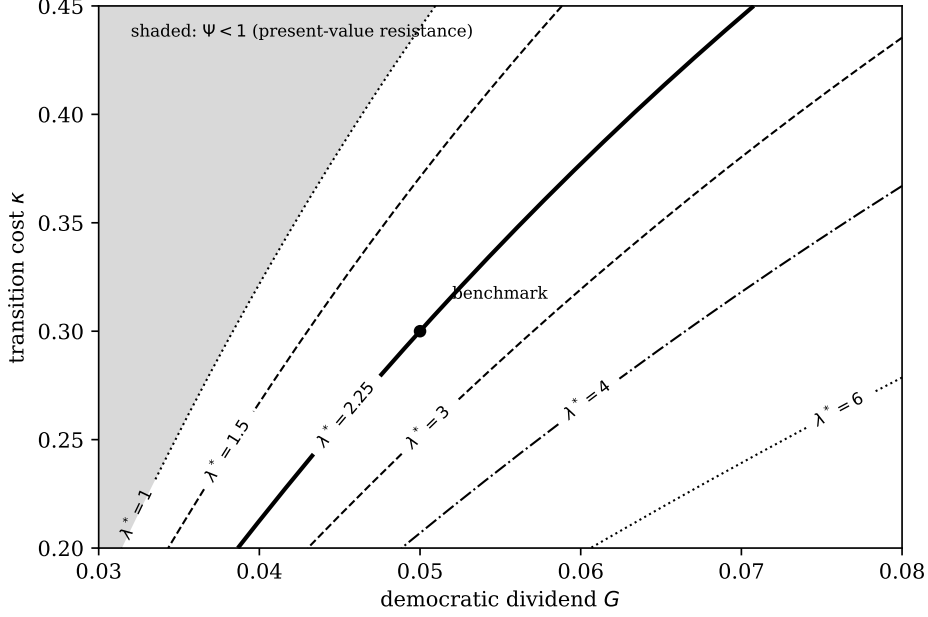


Figure 4: Level sets of the resistance threshold $\lambda^*(G, \kappa)$ at $\tilde{\beta} = 0.80$, $a = 0.02$, $\eta = 1$, so that $\Lambda = (1 + \lambda)/2$. The heavy line is the level set $\lambda^* = 2.25$, on which the illustrative configuration (dot) lies. The shaded region is $\Psi < 1$, where resistance is a present-value prediction (Corollary 1).

unit of present value conceded. At the illustrative configuration it is 1.44, and it exceeds two for elites that are both very loss averse and long-lived. Entrenchment must therefore be substantially cheaper in democratic-quality terms before it is worth using.

8.3 Adaptive Reference, Fragile Restoration, and the Liquidity Bound

Figure 6 illustrates the two extensions. Panel (a) plots $\lambda^*(\chi)$ from Proposition 2 for three values of η : for $\eta = 0.2 < (1 - \tilde{\beta})/\tilde{\beta} = 0.25$ the threshold rises monotonically, while for $\eta = 1$ and $\eta = 2$ it is U-shaped and lies below its fixed-reference value over a wide range of adjustment speeds. Panel (b) plots $\lambda^{**}(\pi)$ from Proposition 3 for three displacement losses under a permanent shortfall ($\psi = 1$); each curve diverges at the critical hazard π^c , beyond which the elite accommodates for every λ . Persistence of the displacement loss is essential to the sign reversal: at $\pi = 0.02$ and $\xi = 0.3$ the threshold is 4.73 when the loss is permanent and 3.34 when it decays at $\psi = 0.7$, and the critical hazard rises from 0.063 to 0.196.

Finally, Table 3 quantifies the liquidity bound of Proposition 5(d) when cash is the only instrument available: the ceiling $\lambda^*(\varsigma/\nu)$ on the loss aversion that cash alone can accommodate. Because the ceiling depends only on ς/ν , a given fiscal and borrowing effort buys less accommodation the more concentrated pre-transition rents are; beyond

Table 1: Relief of the trough required for accommodation, illustrative configuration $(\tilde{\beta}, G, a, \kappa) = (0.80, 0.05, 0.02, 0.30)$

Λ	λ at $\eta = 1$	$\mathcal{L}_\kappa(x^*)$	x^*	x^*/κ
1.75	2.50	0.331	0.018	0.06
2.00	3.00	0.290	0.048	0.16
2.25	3.50	0.258	0.073	0.24
2.50	4.00	0.232	0.093	0.31
3.50	6.00	0.166	0.147	0.49
5.50	10.0	0.105	0.200	0.67

Notes: Accommodation requires $\Lambda \leq \Psi(x)$, so the required dividend statistic is $\Psi = \Lambda$, the implied cost magnitude is $\mathcal{L}_\kappa(x^*) = B/\Lambda$, and $x^* = \kappa - (1 - e^{-\mathcal{L}_\kappa(x^*)})$. The second column translates Λ into λ under $\eta = 1$.

Table 2: Dominance margin of transitional protection over permanent entrenchment, Λ_σ/Λ_q

Effective loss aversion Λ	Effective horizon $\tilde{\beta}$		
	0.70	0.80	0.90
1.25	1.16	1.19	1.22
1.625	1.37	1.44	1.53
2.00	1.54	1.67	1.82
2.50	1.72	1.92	2.17
3.00	1.88	2.14	2.50

Notes: $\Lambda_\sigma/\Lambda_q = \Lambda/[1 + (1 - \tilde{\beta})(\Lambda - 1)]$, which depends on (η, λ) only through Λ . The margin equals one at $\Lambda = 1$.

the ceiling, accommodation requires the non-fiscal instruments of Section 7.3, and if those are unavailable or unacceptable the settlement fails in the sense of Proposition 6.

9 Case Illustrations and Scope Conditions

This section applies the model’s timing convention to three transitions, shows how its primitives can be read off documented instruments, and states the conditions under which the theory does not apply. It is an exercise in illustration, not in identification. No case measures loss aversion, the reference point, or the displacement hazard, and the outcomes cannot discriminate the mechanism from the explanations established in the comparative literature. Section 9.6 says what evidence would discriminate.

9.1 Dating the Decision Window

Definition 1 places $t = 0$ at the first period in which a democratic framework is in force while the elite’s capacity to overturn it remains intact. Table 4 applies it, and dates the close of the *restoration-capacity* window of Definition 2. The behavioral vulnerability

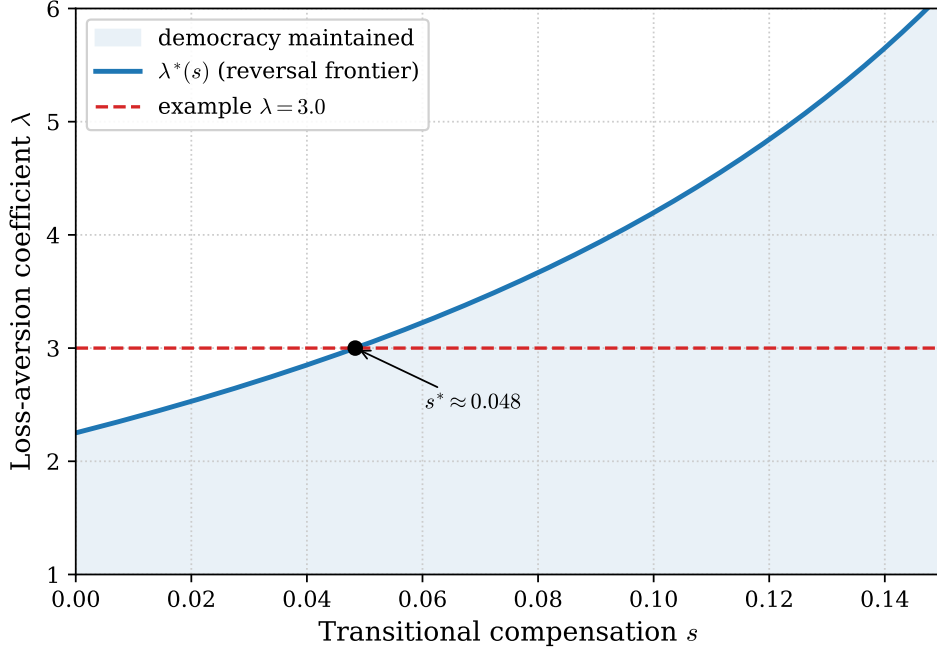


Figure 5: Threshold $\lambda^*(x)$ as a function of period-zero relief $x = s + \sigma$, at $\tilde{\beta} = 0.80$, $G = 0.05$, $a = 0.02$, $\kappa = 0.30$, $\eta = 1$. The elite accommodates in the region below the frontier. An elite with $\lambda = 3.0$ resists at $x = 0$ but accommodates once $x \geq 0.048$.

window is shorter: under the baseline continuation path it is the trough (Lemma 1), and it extends beyond the trough only to the degree that a restoration could recover rents (Remark 2). The convention matters because the model's claims about early and late concessions are claims about position within the *behavioral* window, not about elapsed time since an uprising or about the survival of coercive capacity.

Two implications of the table are worth stating, the first with some care. Spain's legal concessions of October 1977 fall inside the restoration-capacity window, which opens in December 1976, but they do not precede its opening. Lemma 6 requires a credible concession to be vested before the decision it is meant to influence, so the October 1977 instruments bear on the model's decision only under one of two readings. Under the recurrent-veto reading of Lemma 1, the elite's veto remained live through the constituent period, and the trough of a transition that had begun months earlier had not closed, so the amnesty and the pacts were vested while the decision was still open. Under the stricter reading, in which the decisive choice is made once at the opening of the window, the October instruments can have mattered only if they were credibly anticipated in advance—which the negotiated character of the Spanish transition makes plausible but which this paper does not establish. The case is reported here as an illustration of instrument *form*, and the timing claim is deliberately limited to that. Egypt's window is about six months, so evidence drawn from the 2011 uprising describes the anticipation of losses rather than events within the window; that is consistent with κ being defined as

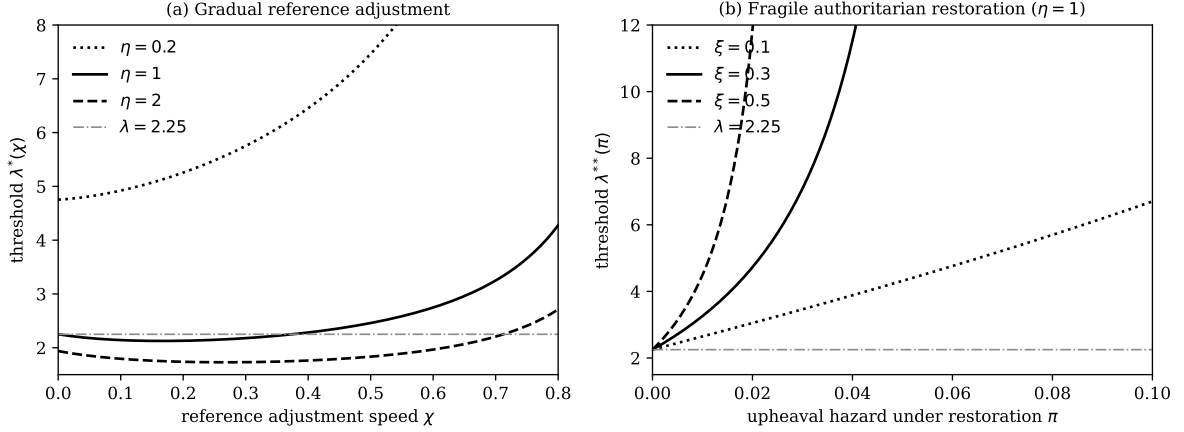


Figure 6: (a) Resistance threshold under gradual reference adjustment, $\lambda^*(\chi)$, for $\eta \in \{0.2, 1, 2\}$. Thresholds are drawn in λ ; at $\eta = 1$ the corresponding effective loss aversion is $\Lambda = (1+\lambda)/2$. (b) Resistance threshold under fragile restoration, $\lambda^{**}(\pi)$, for displacement losses $\xi \in \{0.1, 0.3, 0.5\}$, $\psi = 1$ and $\eta = 1$. Illustrative configuration $\hat{\beta} = 0.80$, $G = 0.05$, $a = 0.02$, $\kappa = 0.30$. The dash-dotted line marks $\lambda = 2.25$.

Table 3: Ceiling on accommodation by cash alone, $\lambda^*(\varsigma/\nu)$, illustrative configuration

Elite share of output ν	Transfer capacity ς (share of output)		
	0.01	0.02	0.03
0.10	4.20	10.01	unbounded
0.20	3.03	4.20	6.13
0.30	2.74	3.36	4.20
0.40	2.61	3.03	3.55

Notes: Entries are the largest λ accommodated by the maximal feasible cash transfer $\bar{s} = \varsigma/\nu$ at $\eta = 1$. “Unbounded” indicates $\varsigma \geq \nu\kappa$. Without compensation the threshold is $\lambda^*(0) = 2.25$.

the anticipated loss, but it should not be presented as evidence about the window itself.

9.2 Spain: Sunset Protections Inside the Window

Spain’s transition proceeded by reform of the existing legality and is the canonical pacted transition (O’Donnell and Schmitter, 1986; Linz and Stepan, 1996). The Amnesty Law of October 1977 extended amnesty to politically motivated acts and, in its Article 2, to offences committed by authorities, officials, and agents of public order, including offences against the exercise of personal rights.² The Moncloa Pacts of 25 October 1977 distributed the costs of macroeconomic adjustment. In the model’s terms the amnesty is a transitional protection: its value falls in the elite’s loss region, it draws on no period-zero liquidity, and it does not entrench a permanent privilege. By Proposition 4 it is the high-leverage

²Ley 46/1977, de 15 de octubre, de Amnistía (BOE núm. 248, 17 October 1977), art. 2(e)–(f).

Table 4: The restoration-capacity window under Definitions 1 and 2

Case	Framework in force ($t = 0$)	Window closes	Concessions to the elite
Spain	Political Reform Law approved by referendum, 18 December 1976; first free elections, June 1977	Constitution of December 1978; residual coup capacity into the early 1980s	Amnesty Law, 15 October 1977; Moncloa Pacts, 25 October 1977 — inside the window
Tunisia	Interim framework, constituent law of 16 December 2011; permanent constitution, 26 January 2014	Uncertain, and probably well before the democratic episode ended: coercive capacity was held by an actor other than the rent-exposed elite (Section 9.3)	Administrative reconciliation law, adopted 13 September 2017 — after the behavioral window closed
Egypt	Constitution signed into force, 26 December 2012 (interim framework from March 2011)	Military removal of the president, 3 July 2013	None

instrument, and by Proposition 5 it is the one a constitutional designer should use first.

9.3 Tunisia: Concentrated Rents, Late Protection, Separated Elites

Firms confiscated from the Ben Ali family after January 2011 accounted for 0.2 per cent of private firms but 5 per cent of private-sector output and 16 per cent of net private-sector profits in 2010 (Rijkers et al., 2014), and connected firms were concentrated in sectors subject to authorisation requirements and restrictions on foreign entry (Rijkers et al., 2017). Both the anticipated loss κ for the connected elite and the rent-concentration index ν were therefore high. By Proposition 5(d), high ν does not make cash less effective—cash and transitional protection have identical leverage—but it makes cash the most fiscally costly trough instrument and the most tightly constrained, since the bound ς/ν tightens as rents concentrate.

Protection came late and narrow. A reconciliation bill covering businessmen, announced in March 2015 and sent to parliament in July 2015, was reduced after sustained opposition to a law covering only officials who had not personally benefited, adopted on 13 September 2017.³ By Proposition 1 and Lemma 1, a protection enacted after the

³Loi organique n° 2017-62 du 24 octobre 2017, relative à la réconciliation dans le domaine administratif.

behavioral vulnerability window has closed lacks the loss-region leverage of a trough concession: a recipient no longer below its benchmark values it at $1 + \eta$ rather than $1 + \eta\lambda$, and after the trough the mechanism no longer generates resistance for the loss weight to amplify. Such a protection may still have political value, but not through the channel studied here. That is the honest reading of 2017, and it is less flattering to the model as an account of 2011–2021: what the model says about Tunisia is confined to the early transition, and the 2017 episode is evidence on the *form* the instrument took—narrow, official-only, and late—not on the behavioral threshold.

The restoration-capacity window is harder to date than the democratic episode, and the table does not identify the two. Democracy survived until July 2021, but that is not evidence that the rent-exposed Ben Ali business elite retained the capacity to restore the prior order until then, and the rupture of 25 July 2021 was carried out by an elected president rather than by that elite. Under the conjunction condition discussed in the next paragraph—resistance requires that the group exposed to the loss also hold coercive capacity—Tunisia is a case in which the capacity window for the relevant elite was probably short, which is itself part of the explanation of the decade that followed.

Tunisia also illustrates a limitation of the representative-elite formulation that matters for the comparison with Egypt. The group bearing the transition loss—the connected business network—was not the group holding the coercive capacity that a restoration requires; the military had been kept out of politics under the prior regime (Grewal, 2023; Stepan and Linz, 2013). In the model’s terms, resistance requires the *conjunction* of exposure to κ and possession of φ , and in Tunisia the two were separated: the hazard π facing any attempted restoration was high, which by Proposition 3 raises the threshold. The rupture of 25 July 2021 was carried out by an elected president, not by the pre-2011 economic elite; by Lemma 1 it lies outside the mechanism, which operates in the trough and through the elite that bears it.

9.4 Egypt: The Conjunction Satisfied

Egypt presents the configuration Tunisia lacked. The Supreme Council of the Armed Forces, which took power on Mubarak’s resignation on 11 February 2011, was simultaneously the principal coercive institution and a major economic proprietor (Sayigh, 2019); connected rents were substantial (Chekir and Diwan, 2014); and protests were associated with lower valuations of firms connected to the group in power relative to unconnected firms (Acemoglu et al., 2018), which is consistent with an anticipated $\kappa > 0$ for connected owners. Exposure to the loss and coercive capacity resided in the same actor, and the coercive apparatus had both the capacity and the will to hold power (Bellin, 2012), so φ was high and π low.

The concessions the 2012 constitution did make were of the low-leverage kind. It

preserved the military’s jurisdiction over civilians for offences harming the armed forces rather than relieving the economic loss.⁴ In the model’s terms this is entrenchment, $q > 0$: by (22) it does relieve the period-zero trough, since the premium enters $t = 0$ as well, but it protects prerogatives in every period and therefore places only the fraction $1 - \tilde{\beta}$ of its present value in the loss region, with the remainder delivered after the elite has left it. Per unit conceded it is the low-leverage instrument of Proposition 4, not an instrument that misses the trough. The window closed with the removal of the president on 3 July 2013, and Egypt’s polyarchy score had never crossed the descriptive threshold of 0.5 (Figure 2).

9.5 Scope Conditions, and Libya

The theory applies where three conditions hold: an identifiable incumbent elite exists; that elite retains a feasible option of authoritarian restoration; and the group exposed to the transition loss overlaps with the group holding coercive capacity. Libya after 2011 satisfies none of them. The Political Isolation Law, passed by the General National Congress on 5 May 2013 while armed groups besieged government ministries, barred those who had held a wide range of public, military, security, and university posts between 1 September 1969 and 23 October 2011 from office for ten years,⁵ which in the model’s terms is a *negative* concession: it deepens the loss instead of relieving it. But no unified coercive apparatus existed that could have restored the prior order (Chivvis, 2014), and what followed was fragmentation and armed conflict from 2014 (Wehrey, 2018). When restoration is not on the menu, the elite’s outside option is contestation, which the model does not represent. Libya is reported here as a boundary case that marks where the theory stops, not as a fourth observation.

9.6 What Would Discriminate

The three illustrations are consistent with the model, and consistency is not discrimination. Military legacies (Grewal, 2023), the coercive robustness of incumbent regimes (Bellin, 2012), and commitment problems (Acemoglu and Robinson, 2001) account for parts of the same pattern, and the classification of latent primitives is partly informed by the outcomes it is then said to organise. Nothing in this section escapes that circularity.

Proposition 1 indicates the way out. The behavioral mechanism predicts that an otherwise identical concession—same legal form, same date, same credibility—has a discontinuously larger effect on the willingness to accommodate when the recipient has been placed below its pre-transition benchmark than when it has not. Within-transition variation of this kind is documented: confiscation lists, exclusion decrees, and licence withdrawals record who was placed below the benchmark, while the reconciliation procedures

⁴Constitution of Egypt (2012), art. 198.

⁵Law No. 13 of 2013 on Political and Administrative Isolation, issued 8 May 2013.

that followed applied to broad categories of actors. A design that compares responses to a common procedure across groups that differ in prior exposure, holding form and date fixed, would test the mechanism rather than illustrate it.

Three things such a design must specify. First, the dependent variable, which must be an observable act of accommodation or obstruction: take-up of a reconciliation procedure by those eligible for it; participation in or public endorsement of a negotiated pact; litigation or constitutional challenge against transitional legislation; participation in, or public support for, a coup attempt among officer cohorts differing in exposure; or, for connected firms, the market response to the announcement of a concession, which is measurable in the way [Acemoglu et al. \(2018\)](#) measure the response to protest. Second, the location of the benchmark, from records of what was actually taken: confiscation lists, exclusion decrees, revoked licences, dismissals. Third, an explicit treatment of measurement error, which matters more here than usual: because the prediction is a jump exactly at the benchmark, error in locating it attenuates the discontinuity, so the test is conservative and requires a defensible bandwidth around the threshold and a placebo check at points where no entitlement boundary lies. The design also requires a setting in which no legal or eligibility cut-off coincides with the benchmark, for the reason given after Proposition 1. This is demanding, and it is the natural next paper rather than a robustness exercise in this one.

10 Conclusion

This paper asks when the timing and form of constitutional concessions matter beyond their present value. The answer it gives rests on a single behavioral premise: that an incumbent elite evaluates a democratic transition against the position the prior order guaranteed it, and weights losses relative to that benchmark more heavily than gains. Under that premise the elite accommodates a transition if and only if its effective loss aversion $\Lambda = (1 + \eta\lambda)/(1 + \eta)$ does not exceed the present value of the democratic dividend per unit of transition cost, and the behavioral content of the mechanism is confined to configurations in which the present-value comparison favours accommodation.

Four results follow. Concessions are worth $(1 + \eta\lambda)$ per unit below the reference and $(1 + \eta)$ above it, so their effectiveness jumps at the benchmark; that discontinuity, not the mere advantage of early concessions, is what separates the mechanism from the class of smooth reference-independent accounts—commitment, liquidity, and horizon explanations all favour front-loading, and none of them produces a jump at the recipient’s inherited position. Gradual adjustment of the reference does not dissolve the mechanism and, when $\eta > (1 - \tilde{\beta})/\tilde{\beta}$, moderate adjustment strengthens it. When restored autocracy is itself fragile the threshold rises, and beyond a critical displacement hazard—which exists only if the loss a failed restoration would inflict is both large and persistent—loss

aversion changes sign and favours accommodation, so behavioral resistance is concentrated where restoration is secure. Finally, because the government cannot commit to pay after the elite's veto has lapsed, credible concessions must be vested at the transition or entrenched in the constitution, and among these a concession concentrated in the trough relaxes the accommodation constraint by the factor $\Lambda/[1 + (1 - \tilde{\beta})(\Lambda - 1)]$ relative to a recurring privilege of equal present value, which is 1.44 at the illustrative configuration. A designer acting for citizens should therefore concede in the trough, through whichever trough instrument citizens price more cheaply—the choice between cash and transitional protection is a price comparison, not a behavioral one—and turn to recurring privilege last. Where the cheapest settlement that secures accommodation costs citizens more than democracy is worth to them, no settlement in the admissible instrument set satisfies both sides and the transition fails. That statement locates a bargaining frontier under the model's reduced-form citizen objective; it is not a constitutional norm.

The limits of the argument should be stated as plainly as the results. The status-quo reference is a primitive, not a derivation, and under an expectations-based reference the mechanism does not operate; the numerical configurations are illustrations rather than calibrations, and the laboratory estimate of loss aversion corresponds to Λ rather than to λ ; the model treats democracy, once accommodated, as absorbing, and the window of vulnerability it identifies is the transition trough, so late breakdowns fall outside it; and the case illustrations establish compatibility, not discrimination. What the paper offers to the constitutional literature is a formal reason why transitional settlements concentrate protection at the moment of transition rather than spreading it over the life of the constitution, and a prediction sharp enough to be refuted.

A Proofs

Proof of Lemma 2

Lifetime value under regime ρ is $V(\rho) = \sum_{t \geq 0} \tilde{\beta}^t [\log c_t^\rho + \eta n(c_t^\rho \mid c_{-1})]$. I compute the differential term by term.

Intrinsic component. The discounted sum of $\log c_t^\rho$ shares the common term $\sum_{t \geq 0} \tilde{\beta}^t \log c_{-1} = \log c_{-1} / (1 - \tilde{\beta})$, which cancels in the differential. Under democracy the remaining growth contribution is $\log(1 - \kappa)$ at $t = 0$ and $\sum_{t \geq 1} \tilde{\beta}^t t \log(1 + G)$ thereafter; using $\sum_{t \geq 1} t \tilde{\beta}^t = \tilde{\beta} / (1 - \tilde{\beta})^2$,

$$\sum_{t \geq 0} \tilde{\beta}^t \log c_t^D - \frac{\log c_{-1}}{1 - \tilde{\beta}} = \log(1 - \kappa) + \frac{\tilde{\beta} \log(1 + G)}{(1 - \tilde{\beta})^2}.$$

Under authoritarian restoration the growth contribution is $\sum_{t \geq 1} \tilde{\beta}^t t \log(1 + a) = \tilde{\beta} \log(1 + a) / (1 - \tilde{\beta})^2$ (the $t = 0$ term is zero). Differencing and using $\log(1 - \kappa) = -\mathcal{L}_\kappa$,

$$\Delta_{\text{intr}} = -\mathcal{L}_\kappa + \frac{\tilde{\beta} [\log(1 + G) - \log(1 + a)]}{(1 - \tilde{\beta})^2} = -\mathcal{L}_\kappa + \Psi \mathcal{L}_\kappa,$$

where the last equality uses the definition (10) of Ψ .

Gain-loss component. Under democracy, $t = 0$ is in the loss region with deviation $\log c_0^D - \log c_{-1} = \log(1 - \kappa) = -\mathcal{L}_\kappa$, weighted by λ ; periods $t \geq 1$ are in the gain region with deviation $t \log(1 + G)$, weight one. Hence

$$\sum_{t \geq 0} \tilde{\beta}^t n(c_t^D \mid c_{-1}) = -\lambda \mathcal{L}_\kappa + \frac{\tilde{\beta} \log(1 + G)}{(1 - \tilde{\beta})^2}.$$

Under authoritarian restoration, $t = 0$ is exactly at the reference ($n = 0$) and $t \geq 1$ are gains, giving $\tilde{\beta} \log(1 + a) / (1 - \tilde{\beta})^2$. The differential is

$$\Delta_{\text{gl}} = -\lambda \mathcal{L}_\kappa + \frac{\tilde{\beta} [\log(1 + G) - \log(1 + a)]}{(1 - \tilde{\beta})^2} = -\lambda \mathcal{L}_\kappa + \Psi \mathcal{L}_\kappa.$$

Combination. The value differential is $\Delta_{\text{intr}} + \eta \Delta_{\text{gl}}$:

$$V(D) - V(N) = (-\mathcal{L}_\kappa + \Psi \mathcal{L}_\kappa) + \eta (-\lambda \mathcal{L}_\kappa + \Psi \mathcal{L}_\kappa) = \mathcal{L}_\kappa [(1 + \eta) \Psi - (1 + \eta \lambda)],$$

which is (9). □

Proof of Remark 5

Write $\lambda^* = \Psi + (\Psi - 1)/\eta$, so $\partial\lambda^*/\partial\Psi = (1 + \eta)/\eta > 0$ and, for any structural parameter $X \in \{\kappa, G, a, \tilde{\beta}\}$, $\text{sign}(\partial\lambda^*/\partial X) = \text{sign}(\partial\Psi/\partial X)$. From (10),

$$\Psi = \frac{\tilde{\beta} [\log(1 + G) - \log(1 + a)]}{(1 - \tilde{\beta})^2 \mathcal{L}_\kappa}, \quad \mathcal{L}_\kappa = -\log(1 - \kappa).$$

Since $\mathcal{L}_\kappa$ is strictly increasing in κ and enters the denominator, $\partial\Psi/\partial\kappa < 0$. The numerator factor $\log(1 + G) - \log(1 + a)$ is increasing in G and decreasing in a , giving $\partial\Psi/\partial G > 0$ and $\partial\Psi/\partial a < 0$. The factor $\tilde{\beta}/(1 - \tilde{\beta})^2$ has derivative $(1 + \tilde{\beta})/(1 - \tilde{\beta})^3 > 0$, so $\partial\Psi/\partial\tilde{\beta} > 0$. For the exit hazard, $\tilde{\beta} = \beta(1 - \delta)$ gives $\partial\tilde{\beta}/\partial\delta = -\beta < 0$, hence $\partial\lambda^*/\partial\delta = (\partial\lambda^*/\partial\tilde{\beta})(\partial\tilde{\beta}/\partial\delta) < 0$. Finally, $\partial\lambda^*/\partial\eta = -(\Psi - 1)/\eta^2 < 0$ whenever $\Psi > 1$. Since the accommodation set is $\{\Lambda \leq \Psi\}$ by Remark 4, the primitive comparative statics are those of Ψ ; the derivative with respect to η holds λ rather than Λ fixed and is reported for completeness. $\square$

Proof of Proposition 2

Normalise $\ell_t \equiv \log c_t - \log c_{-1}$ and $\rho_t \equiv \log r_t - \log c_{-1}$, so that $\rho_{-1} = 0$ and (7) reads $\rho_t = (1 - \chi)\rho_{t-1} + \chi\ell_t$. The gain-loss argument is then $\ell_t - \rho_t = (1 - \chi)(\ell_t - \rho_{t-1})$.

Regions. Under resistance $\ell_t = t \log(1 + a)$ is non-decreasing, and ρ_{t-1} is a weighted average, with weights summing to at most one, of 0 and $\ell_0, \dots, \ell_{t-1}$, all of which are bounded above by ℓ_t ; hence the elite is never in the loss region. Under accommodation $\ell_0 = -\mathcal{L}_\kappa < 0 = \rho_{-1}$, so $t = 0$ is a loss; for $t \geq 1$, $\ell_t = t \log(1 + G)$ strictly exceeds every element of $\{0, \ell_0, \dots, \ell_{t-1}\}$, so the elite is in the gain region.

Differential. Let $d_t \equiv \ell_t^D - \ell_t^N$, so that $d_0 = -\mathcal{L}_\kappa$ and $d_t = t[\log(1 + G) - \log(1 + a)]$ for $t \geq 1$, and let ρ_t^d be the reference differential, which by linearity satisfies $\rho_t^d = (1 - \chi)\rho_{t-1}^d + \chi d_t$ with $\rho_{-1}^d = 0$. The intrinsic component is unaffected by the reference and equals $\sum_{t \geq 0} \tilde{\beta}^t d_t = B - \mathcal{L}_\kappa$. Since the loss weight applies only at $t = 0$ under accommodation, the gain-loss differential is

$$(1 - \chi) \left[-\lambda \mathcal{L}_\kappa + S \right], \quad S \equiv \sum_{t \geq 1} \tilde{\beta}^t (d_t - \rho_{t-1}^d).$$

Now $\sum_{t \geq 1} \tilde{\beta}^t d_t = B$ and, writing $\rho_t^d = \chi \sum_{j=0}^t (1 - \chi)^{t-j} d_j$,

$$\sum_{t \geq 1} \tilde{\beta}^t \rho_{t-1}^d = \tilde{\beta} \sum_{t \geq 0} \tilde{\beta}^t \rho_t^d = \frac{\tilde{\beta} \chi}{1 - \tilde{\beta}(1 - \chi)} \sum_{j \geq 0} \tilde{\beta}^j d_j = \frac{\tilde{\beta} \chi}{1 - \tilde{\beta} + \tilde{\beta} \chi} (B - \mathcal{L}_\kappa).$$

Hence, with $k(\chi) \equiv \tilde{\beta}\chi/(1 - \tilde{\beta} + \tilde{\beta}\chi)$ and $B = \Psi\mathcal{L}_\kappa$,

$$V(\mathbf{D}) - V(\mathbf{N}) = \mathcal{L}_\kappa \left\{ (\Psi - 1) + \eta(1 - \chi) [\Psi - k(\chi)(\Psi - 1) - \lambda] \right\}.$$

For $\chi < 1$ this is non-negative if and only if $\lambda \leq \Psi - k(\chi)(\Psi - 1) + (\Psi - 1)/[\eta(1 - \chi)]$, which is (13) and establishes part (a).

Part (b). With $k'(\chi) = \tilde{\beta}(1 - \tilde{\beta})/(1 - \tilde{\beta} + \tilde{\beta}\chi)^2 > 0$ and $k''(\chi) = -2\tilde{\beta}^2(1 - \tilde{\beta})/(1 - \tilde{\beta} + \tilde{\beta}\chi)^3 < 0$, the term $-k(\chi)(\Psi - 1)$ is strictly convex when $\Psi > 1$, and $(\Psi - 1)/[\eta(1 - \chi)]$ is strictly convex and diverges as $\chi \rightarrow 1$. Their sum is strictly convex and diverges.

Part (c). Differentiating,

$$\frac{\partial \lambda^*(\chi)}{\partial \chi} = (\Psi - 1) \left[\frac{1}{\eta(1 - \chi)^2} - \frac{\tilde{\beta}(1 - \tilde{\beta})}{(1 - \tilde{\beta} + \tilde{\beta}\chi)^2} \right],$$

so that

$$\left. \frac{\partial \lambda^*(\chi)}{\partial \chi} \right|_{\chi=0} = (\Psi - 1) \left[\frac{1}{\eta} - \frac{\tilde{\beta}}{1 - \tilde{\beta}} \right].$$

By strict convexity the derivative is strictly increasing in χ . If $\eta \leq (1 - \tilde{\beta})/\tilde{\beta}$, it is non-negative at $\chi = 0$ and hence strictly positive on $(0, 1)$. Otherwise it is negative at $\chi = 0$ and, since λ^* diverges, has a unique zero, which solves $\eta\tilde{\beta}(1 - \tilde{\beta})(1 - \chi)^2 = (1 - \tilde{\beta} + \tilde{\beta}\chi)^2$; taking positive square roots gives (14), which lies in $(0, 1)$ precisely when $\eta\tilde{\beta}(1 - \tilde{\beta}) > (1 - \tilde{\beta})^2$. For $\chi > 0$, $\lambda^*(\chi) = \lambda^*(0)$ if and only if $k(\chi)(\Psi - 1) = (\Psi - 1)\chi/[\eta(1 - \chi)]$, that is, $\eta\tilde{\beta}(1 - \chi) = 1 - \tilde{\beta} + \tilde{\beta}\chi$, whose solution is (15); it lies in $(0, 1)$ under the same condition. By strict convexity, $\lambda^*(\chi) < \lambda^*(0)$ on $(0, \hat{\chi})$.

Part (d). The elite resists at χ if and only if $\lambda^*(\chi) < \lambda$. The strict sublevel sets of a continuous strictly convex function are intervals. If $\lambda > \lambda^*(0)$, the interval contains 0, and since λ^* diverges it has an upper endpoint $\bar{\chi}(\lambda) < 1$, defined by $\lambda^*(\bar{\chi}) = \lambda$ on the increasing branch of λ^* ; hence $\bar{\chi}$ is strictly increasing in λ . Because $\lambda^*(\hat{\chi}) = \lambda^*(0) < \lambda$, $\bar{\chi}(\lambda) > \hat{\chi}$. $\square$

Proof of Lemma 3

Normalise $\log c_{-1} = 0$; the common term $\log c_{-1}/(1 - \tilde{\beta})$ cancels because both streams are discounted at $\tilde{\beta}$ over the same infinite horizon. Under accommodation, the computation of Lemma 2 gives $V(\mathbf{D}) = -(1 + \eta\lambda)\mathcal{L}_\kappa + (1 + \eta)\tilde{\beta} \log(1 + G)/(1 - \tilde{\beta})^2$. Under resistance, the restored regime survives through period t with probability $(1 - \pi)^t$. In a surviving period the elite consumes $c_{-1}(1 + a)^t \geq c_{-1}$ and has flow utility $(1 + \eta)t \log(1 + a)$; k periods after displacement its log consumption falls short of the reference by $\psi^k \Xi$, so its flow utility is $-(1 + \eta\lambda)\psi^k \Xi$. Displacement occurs at date $d \geq 1$ with probability

$(1 - \pi)^{d-1}\pi$, and $\sum_{t \geq d} \tilde{\beta}^t \psi^{t-d} = \tilde{\beta}^d / (1 - \tilde{\beta}\psi)$, so the discounted expected shortfall is

$$\Xi \sum_{d \geq 1} (1 - \pi)^{d-1} \pi \frac{\tilde{\beta}^d}{1 - \tilde{\beta}\psi} = \frac{\tilde{\beta} \pi \Xi}{(1 - \tilde{\beta}\psi)(1 - \tilde{\beta} + \tilde{\beta}\pi)} = E_\pi.$$

With $\sum_{t \geq 0} \tilde{\beta}_\pi^t t = \tilde{\beta}_\pi / (1 - \tilde{\beta}_\pi)^2$ for the surviving component, differencing gives (17). $\square$

Proof of Proposition 3

Because $x \mapsto x/(1 - x)^2$ is strictly increasing on $(0, 1)$ and $\tilde{\beta}_\pi$ is decreasing in π , B_π is non-decreasing in π (strictly if $a > 0$), independent of ξ , and $B_\pi \geq B_0 = B > 0$. The expected displacement loss E_π is zero at $\pi = 0$, strictly increasing in π because $\pi/(1 - \tilde{\beta} + \tilde{\beta}\pi)$ is, and, for $\pi > 0$, strictly increasing in Ξ and hence in ξ , and in ψ because $1/(1 - \tilde{\beta}\psi)$ is. Hence $M_\pi = \mathcal{L}_\kappa - E_\pi$ is strictly decreasing in π , and $M_\pi = 0$ solves $\tilde{\beta}\pi\Xi = (1 - \tilde{\beta}\psi)(1 - \tilde{\beta} + \tilde{\beta}\pi)\mathcal{L}_\kappa$, which gives (18); a solution in $(0, 1)$ exists if and only if $\tilde{\beta}\Xi > (1 - \tilde{\beta}\psi)\mathcal{L}_\kappa$.

Part (a). If $M_\pi > 0$, (17) is non-negative if and only if $1 + \eta\lambda \leq (1 + \eta)B_\pi/M_\pi$, which is (19). Since $B_\pi \geq B$ and $M_\pi \leq \mathcal{L}_\kappa$, with at least one strict inequality when $\pi > 0$, $\Psi_\pi \geq \Psi$ with equality if and only if $\pi = 0$, and Ψ_π is strictly increasing in π and, for $\pi > 0$, in ξ . As λ^{**} is strictly increasing in Ψ_π , the claims follow.

Part (b). If $M_\pi \leq 0$, then $V(D) - V(N) = (1 + \eta)B_\pi - (1 + \eta\lambda)M_\pi \geq (1 + \eta)B_\pi > 0$ for all $\lambda \geq 1$, and $\partial[V(D) - V(N)]/\partial\lambda = -\eta M_\pi \geq 0$. $\square$

Proof of Remark 6

With $\kappa(s) = \kappa - s$, the effective magnitude is $\mathcal{L}_\kappa(s) = -\log(1 - \kappa + s)$, which is strictly decreasing in s on $[0, \kappa)$ because $1 - \kappa + s$ is increasing toward 1. From (21), $\Psi(s)$ is inversely proportional to $\mathcal{L}_\kappa(s)$ with a positive numerator, so $\Psi(s)$ is strictly increasing in s . Since $\lambda^*(s) = \Psi(s) + (\Psi(s) - 1)/\eta$ is strictly increasing in $\Psi(s)$, it is strictly increasing in s . Explicitly,

$$\frac{\partial \lambda^*(s)}{\partial s} = \frac{1 + \eta}{\eta} \frac{\partial \Psi(s)}{\partial s} = \frac{1 + \eta}{\eta} \frac{\tilde{\beta}[\log(1 + G) - \log(1 + a)]}{(1 - \tilde{\beta})^2 \mathcal{L}_\kappa(s)^2} \frac{1}{1 - \kappa + s} > 0,$$

using $\partial \mathcal{L}_\kappa(s)/\partial s = -1/(1 - \kappa + s)$. $\square$

Proof of Lemma 1

Normalise $\log c_{-1} = 0$, so that the accommodation path is $\ell_0 = -\mathcal{L}_\kappa$ and $\ell_t = t \log(1 + G)$ for $t \geq 1$, and restoration at τ gives $\ell_t = t \log(1 + a)$ for $t \geq \tau$ and the accommodation path before τ . Fix $\tau \geq T_0$, so that the trough is over by τ . The two continuations coincide for $t < \tau$. For $t \geq \tau$ both the accommodation path and the authoritarian counterfactual

are weakly above the reference, so the elite is in the gain region under either continuation and flow utility is $(1 + \eta)\ell_t$; note that this holds for $a = 0$ as well, in which case the restored path sits exactly at the reference and the gain-loss term is zero. Hence

$$V(\text{never}) - V(\text{restore at } \tau) = (1 + \eta) \left[\log(1 + G) - \log(1 + a) \right] \sum_{t \geq \tau} \tilde{\beta}^t t > 0,$$

because $G > a$, which proves (a). With $T_0 = 1$ the only remaining date is $\tau = 0$, which gives (b); with $T_0 > 1$ the remaining dates are $\tau \in \{0, \dots, T_0 - 1\}$, a finite set, which gives (c). $\square$

Proof of Proposition 1

(a) Lifetime value is additively separable across dates, and the period-zero contribution to $V(\mathbf{D})$ is $\ell_0 + \eta n(\ell_0)$ with $n(\ell_0) = \lambda \ell_0$ for $\ell_0 < 0$ and $n(\ell_0) = \ell_0$ for $\ell_0 \geq 0$. A perturbation $d\ell_0 = \varepsilon$ therefore changes $V(\mathbf{D}) - V(\mathbf{N})$ by $(1 + \eta\lambda)\varepsilon$ in the loss region and by $(1 + \eta)\varepsilon$ in the gain region; no other date is affected. (b) Group 1 has $\ell_0 = -\mathcal{L}_{\kappa_1} < 0$ and group 2 has $\ell_0 \geq 0$, so the ratio of the two effects is $(1 + \eta\lambda)/(1 + \eta) = \Lambda$. (c) Let preferences be reference-independent, U smooth and increasing, with date weights $\omega_t \geq 0$ encoding discounting, political survival, or the credibility of a promise made for date t . The period-zero contribution is $\omega_0 U(c_0)$, whose derivative with respect to $\log c_0$ is $\omega_0 U'(c_0) c_0$, a continuous function of c_0 and hence of κ . For $U = \log$ it equals ω_0 and is independent of κ , so the ratio in (b) is one; for $U(c) = c^{1-\gamma}/(1-\gamma)$ it equals $\omega_0 c_0^{1-\gamma}$, so the ratio for two groups differing only in κ is $(1 - \kappa_1)^{1-\gamma}$, continuous in κ_1 and equal to one at $\kappa_1 = 0$. In every such case the marginal value varies continuously in the recipient's position, whereas by (a) reference dependence makes it jump at the benchmark. The same continuity holds for reference-dependent preferences with $\lambda = 1$. $\square$

Proof of Lemma 4

Normalise $\log c_{-1} = 0$ and write $m = \log(1+q)$. Under the settlement, $\ell_0 = -\mathcal{L}_\kappa(s, \sigma) + m$ and $\ell_t = t \log(1 + G(q)) + m > 0$ for $t \geq 1$, so every period $t \geq 1$ is in the gain region with flow utility $(1 + \eta)\ell_t$. Under restoration $V(\mathbf{N}) = (1 + \eta) \log(1 + a) \sum_{t \geq 1} \tilde{\beta}^t t$. Using $\sum_{t \geq 1} \tilde{\beta}^t t = \tilde{\beta}/(1 - \tilde{\beta})^2$ and $\sum_{t \geq 1} \tilde{\beta}^t = \tilde{\beta}/(1 - \tilde{\beta})$, the $t \geq 1$ contribution to the differential is $(1 + \eta)[B(q) + \tilde{\beta}m/(1 - \tilde{\beta})]$ in either case. At $t = 0$ the elite is in the loss region if $\ell_0 < 0$, that is $m < \mathcal{L}_\kappa(s, \sigma)$, contributing $-(1 + \eta\lambda)[\mathcal{L}_\kappa(s, \sigma) - m]$, and in the gain region otherwise, contributing $(1 + \eta)[m - \mathcal{L}_\kappa(s, \sigma)]$, while the restoration path is exactly at the reference at $t = 0$ and contributes nothing. Collecting terms gives the two branches of (22). $\square$

Proof of Lemma 5

On the branch $m \geq \mathcal{L}_\kappa(s, \sigma)$, (22) gives

$$V(D) - V(N) = (1 + \eta) \left[\underbrace{m - \mathcal{L}_\kappa(s, \sigma)}_{\geq 0} + \underbrace{B(q)}_{> 0} + \underbrace{\frac{\tilde{\beta}m}{1 - \tilde{\beta}}}_{\geq 0} \right] > 0,$$

where $B(q) > 0$ because $G(q) > a$ for every admissible q . The expression does not involve λ , so accommodation is strict for every $\lambda \geq 1$ and $\eta \geq 0$. Hence the accommodation constraint $V(D) - V(N) \geq 0$ cannot hold with equality on that branch, and any settlement at which it binds satisfies $m < \mathcal{L}_\kappa(s, \sigma)$ strictly. $\square$

Proof of Lemma 6

Consider any history in which the elite has accommodated, and any date $t \geq 1$. By Lemma 1 the elite would not restore authoritarian rule after the trough even if it could, so no payment at $t \geq 1$ can alter its behaviour: the payment s_t affects no subsequent choice by either player. By (23) the government's objective is strictly decreasing in s_t at rate $\nu > 0$, so $s_t = 0$ is its strictly dominant choice in that subgame, independently of continuation play. The same argument applies at every subgame beginning at any $t \geq 1$ and on every path—no terminal date is required—so every subgame-perfect equilibrium has $s_t = 0$ for all $t \geq 1$, and the elite values an announced schedule $\{s_t\}_{t \geq 1}$ at zero. Stage-(3) choices do not include (σ, q) , which are enacted at stage (1) and can be altered only through the amendment procedure, so their value to the elite is unaffected. $\square$

Proof of Proposition 4

By (22) the value differential is linear in $\mathcal{L}_\kappa(s, \sigma)$ and in m . A concession that reduces $\mathcal{L}_\kappa$ by P raises it by $(1 + \eta\lambda)P$, and the present value delivered to the elite is P itself, since the relief accrues at $t = 0$; hence $\Lambda_s = \Lambda_\sigma = 1 + \eta\lambda$. Entrenchment m raises the differential by $(1 + \eta\lambda)m + (1 + \eta)m\tilde{\beta}/(1 - \tilde{\beta})$ and delivers present value $m \sum_{t \geq 0} \tilde{\beta}^t = m/(1 - \tilde{\beta})$, so

$$\Lambda_q = (1 - \tilde{\beta}) \left[(1 + \eta\lambda) + \frac{(1 + \eta)\tilde{\beta}}{1 - \tilde{\beta}} \right] = (1 - \tilde{\beta})(1 + \eta\lambda) + \tilde{\beta}(1 + \eta) = (1 + \eta) + (1 - \tilde{\beta})\eta(\lambda - 1).$$

Thus $\Lambda_\sigma - \Lambda_q = \tilde{\beta}\eta(\lambda - 1) \geq 0$, with equality if and only if $\lambda = 1$, since $\tilde{\beta} \in (0, 1)$. If the dividend depends on q , differentiating (22) with respect to q gives the additional term $(1 + \eta)B'(q) dq$, while the present value delivered is $dP_q = dq/[(1 + q)(1 - \tilde{\beta})]$; hence the net leverage is $\Lambda_q^{\text{eff}} = \Lambda_q + (1 + \eta)(1 - \tilde{\beta})(1 + q)B'(q) < \Lambda_q$. Substituting $B'(q) = \tilde{\beta}G'(q)/[(1 + G(q))(1 - \tilde{\beta})^2]$ shows that $\Lambda_q^{\text{eff}} > 0$ if and only if (24) holds, and that

$\Lambda_q^{\text{eff}} \leq 0$ otherwise, in which case entrenchment tightens the accommodation constraint and is never used. Dividing numerator and denominator of Λ_σ/Λ_q by $1 + \eta$ and using $\eta(\lambda - 1)/(1 + \eta) = \Lambda - 1$ gives $\Lambda_\sigma/\Lambda_q = \Lambda/[1 + (1 - \tilde{\beta})(\Lambda - 1)]$, whose derivative with respect to Λ is $\tilde{\beta}/[1 + (1 - \tilde{\beta})(\Lambda - 1)]^2 > 0$ and which is increasing in $\tilde{\beta}$ because the denominator is decreasing in $\tilde{\beta}$. $\square$

Proof of Propositions 5 and 6

Throughout, by Lemma 5 the accommodation constraint binds only where $m < \mathcal{L}_\kappa(s, \sigma)$, so the first branch of (22) applies at any optimum with $D > 0$ and the constraint $m \leq \mathcal{L}_\kappa(s, \sigma)$ is imposed without loss.

(a) W is strictly decreasing in s , in σ , and in q by (23) and $B' < 0$. If $D \leq 0$ the accommodation constraint holds at $(0, 0, 0)$, which is therefore the unique optimum.

(b) Let $D > 0$. W is strictly decreasing in every instrument, and each instrument with positive net leverage relaxes the constraint—cash and transitional protection always, entrenchment when (24) holds; where it fails, $\Lambda_q^{\text{eff}} \leq 0$ and $q = 0$ at any optimum, since $q > 0$ would both lower W and tighten the constraint. The constraint binds at the optimum, because otherwise all instruments could be reduced. Write the constraint as $\mathcal{R}(s, \sigma, q) \geq D$, where $\mathcal{R}$ is the relaxation implied by (22); it is linear in the relief of $\mathcal{L}_\kappa$ and in m , and in q it also carries the dividend term, so that $\partial \mathcal{R} / \partial z_j = \Lambda_j^{\text{eff}} \partial P_j / \partial z_j$ with $\Lambda_s^{\text{eff}} = \Lambda_s$, $\Lambda_\sigma^{\text{eff}} = \Lambda_\sigma$, and Λ_q^{eff} as in Proposition 4. The first-order conditions with multiplier $\mu > 0$ are $\partial W / \partial z_j + \mu \Lambda_j^{\text{eff}} \partial P_j / \partial z_j = 0$ at any interior optimum, that is $p_j / \Lambda_j^{\text{eff}} = \mu$ for every such instrument, where $p_j \equiv -(\partial W / \partial z_j) / (\partial P_j / \partial z_j)$. An instrument with $p_j / \Lambda_j^{\text{eff}} > \mu$ at $z_j = 0$ is not used; one whose bound binds has $p_j / \Lambda_j^{\text{eff}} < \mu$. Since $p_s / p_\sigma = \nu / c_\sigma$ is constant while $\Lambda_s^{\text{eff}} = \Lambda_\sigma^{\text{eff}}$, the ratios of cash and protection are ordered uniformly, so the cheaper instrument is used to its bound before the other is used at all; that comparison is therefore global. For entrenchment, $p_q / \Lambda_q^{\text{eff}}$ may be non-monotone in q under the maintained assumptions, so the first-order conditions characterise a local optimum; if $p_q / \Lambda_q^{\text{eff}}$ is non-decreasing in q , the characterisation is global.

(c) If $p_\sigma = p_q$ then $p_\sigma / \Lambda_\sigma < p_q / \Lambda_q^{\text{eff}}$, since $\Lambda_\sigma > \Lambda_q > \Lambda_q^{\text{eff}}$ for $\lambda > 1$, so by (b) entrenchment is not used until the transitional-protection instrument reaches its bound. Entrenchment is used only if $p_q / p_\sigma \leq \Lambda_q^{\text{eff}} / \Lambda_\sigma$. Gross of the distortion, $\Lambda_\sigma / \Lambda_q$ is increasing in Λ and in $\tilde{\beta}$ by Proposition 4, so the admissible ratio falls. Net of the distortion, write $C \equiv (1 + \eta)(1 - \tilde{\beta})(1 + q)B'(q) < 0$, which does not depend on λ , and

$$\frac{\Lambda_q^{\text{eff}}}{\Lambda_\sigma} = (1 - \tilde{\beta}) + \frac{\tilde{\beta}(1 + \eta) + C}{1 + \eta\lambda}.$$

Differentiating, the ratio is non-increasing in λ if and only if $\tilde{\beta}(1 + \eta) + C \geq 0$, that is $-(1 + q)B'(q) \leq \tilde{\beta}/(1 - \tilde{\beta})$; and it is non-increasing in $\tilde{\beta}$ if $(1 + \eta)[1 + (1 + q)|B'(q)|] \leq 1 + \eta\lambda$,

that is $-(1+q)B'(q) \leq \Lambda - 1$. Condition (25) imposes both, and without it the ratio can increase in λ , so the comparative statics of the admissible price must then be evaluated directly.

(d) $\Lambda_s = \Lambda_\sigma$ by Proposition 4; both instruments are limited by $s + \sigma < \kappa$, and only s faces the fiscal bound ς/ν , which tightens as ν rises.

For Proposition 6, let $\mathcal{C}(D)$ be the value of $\min\{\nu s + c_\sigma \sigma + [B^c(0) - B^c(q)]\}$ subject to $\mathcal{R}(s, \sigma, q) \geq D$, the constraint $m \leq \mathcal{L}_\kappa(s, \sigma)$, and the bounds; $\mathcal{C}$ is non-decreasing in D , because a larger gap shrinks the feasible set, and equals $+\infty$ if the bounds make D unattainable. No curvature is used below. Citizens accept the settlement if and only if $W \geq 0$, that is $\mathcal{C}(D) \leq B^c(0)$, since the cost of entrenchment is already included in $\mathcal{C}$. The gap $D = (1 + \eta\lambda)\mathcal{L}_\kappa - (1 + \eta)B(0)$ is increasing in λ and, because $\mathcal{L}_\kappa$ is increasing in κ and B independent of it, in κ ; the cash bound ς/ν is decreasing in ν and increasing in ς , and the marginal price $p_s = \nu(1 - \kappa + x)$ is increasing in ν ; and the attainable set expands in $\bar{q}$. Holding other primitives fixed, the infeasible region therefore expands in λ , κ , and ν and contracts in ς , $\bar{q}$, and $B^c(0)$. $\square$

Statements and Declarations

Competing interests. The author declares no competing interests.

Funding. No specific funding was received for this research.

Data availability and replication. No new data were generated. The figures in Section 2 use the publicly available V-Dem dataset (version 16). The code that produces every figure, table, and numerical statement in the paper is provided as supplementary material (`make_figures_CPE.py` and `replication.py`).

References

- Acemoglu, D., and Robinson, J. A. (2001). A theory of political transitions. *American Economic Review*, 91(4), 938–963.
- Acemoglu, D., Hassan, T. A., and Tahoun, A. (2018). The power of the street: Evidence from Egypt’s Arab Spring. *Review of Financial Studies*, 31(1), 1–42.
- Albertus, M., and Menaldo, V. (2018). *Authoritarianism and the Elite Origins of Democracy*. Cambridge University Press.
- Alesina, A., and Passarelli, F. (2019). Loss aversion in politics. *American Journal of Political Science*, 63(4), 936–947.
- Bellin, E. (2012). Reconsidering the robustness of authoritarianism in the Middle East: Lessons from the Arab Spring. *Comparative Politics*, 44(2), 127–149.

- Bermeo, N. (2016). On democratic backsliding. *Journal of Democracy*, 27(1), 5–19.
- Besley, T. (2006). *Principled Agents? The Political Economy of Good Government*. Oxford University Press.
- Chekir, H., and Diwan, I. (2014). Crony capitalism in Egypt. *Journal of Globalization and Development*, 5(2), 177–211.
- Chivvis, C. S. (2014). *Toppling Qaddafi: Libya and the Limits of Liberal Intervention*. Cambridge University Press.
- Grewal, S. (2023). *Soldiers of Democracy? Military Legacies and the Arab Spring*. Oxford University Press.
- Haggard, S., and Kaufman, R. (2021). *Backsliding: Democratic Regress in the Contemporary World*. Cambridge University Press.
- Kőszegi, B., and Rabin, M. (2006). A model of reference-dependent preferences. *Quarterly Journal of Economics*, 121(4), 1133–1165.
- Linz, J. J., and Stepan, A. (1996). *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-Communist Europe*. Johns Hopkins University Press.
- Lührmann, A., Tannenberg, M., and Lindberg, S. I. (2018). Regimes of the World (RoW): Opening new avenues for the comparative study of political regimes. *Politics and Governance*, 6(1), 60–77.
- Lührmann, A., and Lindberg, S. I. (2019). A third wave of autocratization is here: what is new about it? *Democratization*, 26(7), 1095–1113.
- O'Donnell, G., and Schmitter, P. C. (1986). *Transitions from Authoritarian Rule: Tentative Conclusions about Uncertain Democracies*. Johns Hopkins University Press.
- Pagel, M. (2017). Expectations-based reference-dependent life-cycle consumption. *Review of Economic Studies*, 84(2), 885–934.
- Passarelli, F., and Tabellini, G. (2017). Emotions and political unrest. *Journal of Political Economy*, 125(3), 903–946.
- Persson, T., and Tabellini, G. (2009). Democratic capital: the nexus of political and economic change. *American Economic Journal: Macroeconomics*, 1(2), 88–126.
- Rijkers, B., Freund, C., and Nucifora, A. (2014). All in the family: State capture in Tunisia. Policy Research Working Paper 6810, World Bank.

- Rijkers, B., Freund, C., and Nucifora, A. (2017). All in the family: State capture in Tunisia. *Journal of Development Economics*, 124, 41–59.
- Sayigh, Y. (2019). *Owners of the Republic: An Anatomy of Egypt’s Military Economy*. Carnegie Middle East Center.
- Stepan, A., and Linz, J. J. (2013). Democratization theory and the “Arab Spring”. *Journal of Democracy*, 24(2), 15–30.
- Tversky, A., and Kahneman, D. (1991). Loss aversion in riskless choice: a reference-dependent model. *Quarterly Journal of Economics*, 106(4), 1039–1061.
- Tversky, A., and Kahneman, D. (1992). Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4), 297–323.
- Coppedge, M., Gerring, J., Knutsen, C. H., Lindberg, S. I., Teorell, J., et al. (2026). V-Dem Country–Year Dataset v16. Varieties of Democracy (V-Dem) Project, University of Gothenburg.
- Wehrey, F. (2018). *The Burning Shores: Inside the Battle for the New Libya*. Farrar, Straus and Giroux.