Splits as Price Wars

Party discipline under imperfect monitoring

Torun Dewan

Preliminary working note. Comments welcome.

Abstract

A party of two factions holds to a common manifesto only by tolerating occasional splits. Each faction privately chooses whether to advocate the agreed line or press its own ideal. That choice is not observed. What is observed is a noisy signal of electoral support, so a fall in support cannot be told apart from a defection. In the trigger-strategy equilibrium the factions unite while support runs high. When it falls they *split* – advocating their own lines, at a cost to the party's standing – even though, on the equilibrium path, neither has defected. Splits are the false alarms a disciplined party must bear. They are costly in the short run, but they deter deviation and sustain unity in the long run. We characterise the scheme and its comparative statics. We then note that the symmetry of a split leaves it open to renegotiation: a party, or a leader, unable to commit to endure one forfeits its discipline. The model gives the party-as-cartel of Cox and McCubbins (2005) the cartel theory its name invokes, the Green and Porter (1984) account of collusion under imperfect monitoring, and with it an account of the on-path splits the cartel thesis omits.

1 Introduction

Krehbiel (1993) asked which party behaviour is significant and found the usual answers wanting: much of what looks like party influence is only the shared preferences of members who would have voted alike in any case. The task is to identify behaviour that is the party's own. Cox and McCubbins (1993, 2005) supply an answer. The majority party acts as a legislative *cartel*, monopolising the agenda to protect a collective good its members value above going their own way: the party's brand, its control of business, its majority. The same form recurs in European politics as the *cartel party* of Katz and Mair (1995), where the collective good is pecuniary – state resources and public funding, allocated in proportion to size, so that to fragment is to forfeit the subvention that scale commands. In both settings the party is a cartel holding a collective good against its members' temptation to defect.

The cartel's logic is a folk theorem: cooperation among self-interested members sustained by the shadow of a valuable joint enterprise. It has not, however, been set down as a repeated game, and it offers no account of the breakdowns observed on the equilibrium path – the rebellions, defections, and splits that occur even in disciplined parties. This paper provides that

formalisation, modelling the cartel as a repeated game of imperfect monitoring, and uses it to develop a theory of party splits.

We model the party as a cartel in this sense. Since Green and Porter (1984) the economics of the cartel has been a theory of cooperation under *imperfect monitoring*, and that is where the on-path breakdowns arise. A party is a coalition of factions that must act as one. To contest an election it offers a single manifesto; to govern it holds a line. Each faction keeps its own ideal, and is tempted, in the daily work of advocacy, to press it. The party prospers only if its factions hold the common line, but holding a line that is not one's own is what each would rather not do. The party cannot see whether a faction has held the line. It observes only a noisy signal of electoral support. So it must discipline the temptation as a cartel disciplines cheating, by punishing the signal.

That punishment is the split, and on the equilibrium path it falls on factions that have not defected. A cartel that cannot observe output triggers price wars on a noisy price. It punishes slumps it knows may be mere bad luck, because to forgive every doubtful slump is to invite cheating. A party triggers splits on a noisy signal of support, for the same reason. The split on the blameless is the mechanism of the scheme rather than a flaw in it. The on-path breakdown that the cartel thesis cannot accommodate becomes the centre of the account. This reverses the usual reading of intra-party division, much as Dewan and Squintani (2016) reversed the reading of factions themselves, and in keeping with a literature that treats intra-party dissent as strategic rather than pathological (Izzo, 2024; Invernizzi and Izzo, 2024). The account rests on five results.

First, discipline cannot dispense with splits. A party that resolved never to split could not deter its factions. If every fall in support were forgiven as bad luck, a faction could press its own line whenever support dipped, confident the dip would be blamed on the electorate and not on it; anticipating that, no faction would hold the common line. Splits must therefore occur on the equilibrium path, set off by weak support rather than by any observed disloyalty. The party acts when it cannot tell a slump from a defection, and it is the readiness to punish the possibly innocent that keeps the genuinely disloyal in check.

Second, the optimal split is rare and severe rather than frequent and mild. Splits are costly, so a party that could design its own discipline would rather they came seldom; but a punishment deters only if it hurts. The two pulls are best reconciled by forgiving widely and striking hard. The party tolerates ordinary falls in support and reserves the split for the worst of them: a regime of frequent, mild splits would be both more costly and less of a deterrent. The party forgives as much as deterrence allows, and no more.

Third, once the punishment is made robust to renegotiation, the split takes the form of a *capture*: the party's line passes to one wing, which governs for a time. A symmetric split, in which both wings abandon the common line and all are worse off, is a punishment the party would want to call off the moment it began, and a punishment everyone wishes to cancel is no

credible threat. A capture is self-enforcing instead. The favoured wing prefers its turn and has no reason to end it, so the punishment carries itself out, and the threat of it is believed.

Fourth, when the factions' electoral fortunes move together, the party can do better than a blameless split. Fortunes that rise and fall together carry a common shock that can be netted out: a slump that hits both wings reflects the national mood, while a slump that hits one alone points to that one. The party can then tell defection from misfortune, identify the deviator, and let the punishment fall on the guilty wing rather than on all. A collective, blameless split is in turn the mark of factions whose fortunes are independent, where no such comparison is available.

Fifth, when the manifesto is itself chosen, discipline pulls it toward the party's centre. The line the factions undertake to defend is not handed to them; the party sets it. A line far from a faction's own ideal is harder to ask it to hold, so the discipline that holds the party together draws the agreed line inward, to where both wings can bear it. Beyond a critical divergence between the wings, no line lies close enough to either to be worth holding: discipline fails, and the party does not merely split for a season but breaks for good.

These results yield several predictions: splits follow bad electoral luck rather than bad faith; divided and hard-to-read parties split more often; and a threshold in the distance between the wings separates the parties that endure from those that break.

On the reading that results, significant party behaviour is not the shared preferences the sceptic feared, nor procedure alone, but discipline held by punishing the signal of disloyalty rather than by observing loyalty itself.[1]

Section 2 sets out the model; Section 3 the equilibrium and Section 4 the optimal scheme, with the comparative statics in Section 5; Sections 6 and 7 take up renegotiation and the form a split takes, Section 8 the attribution of blame when factional fortunes are correlated; Section 9 endogenises the manifesto, and Section 10 draws out the empirical implications. Section 12 places the paper among the literatures it meets.

2 Model

A party comprises two factions $i \in \{L, R\}$ with ideal points $x_L < x_R$. They have agreed a manifesto $m \in [x_L, x_R]$ in the Pareto interval between them. Time is discrete and infinite; factions discount the future at rate $\delta \in (0, 1)$. Each period every faction privately chooses an action $a_i \in \{A, D\}$: to *advocate* the line m, or to *deviate* toward its own ideal x_i.

Two things move with these choices. A faction that deviates pulls the party's effective position toward its ideal, a private policy gain $g > 0$. And division costs votes: writing $d \in$

[1] The paper supersedes an earlier and now-defunct treatment in the author's doctoral work, *The Party's Over* (Dewan, 2002), which set out the idea, the model, and its first comparative statics. The present paper carries that treatment further, with a full set of comparative statics, the renegotiation-proofness of the punishment, and the further extensions developed below.

$\{0, 1, 2\}$ for the number of deviations, electoral support is

$$y = \bar{s} - \kappa\, d + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2),\ \kappa > 0, \tag{1}$$

and each faction earns an office return $\phi\, y$ from the party's electoral standing, $\phi > 0$. A faction's stage payoff is thus $\phi\, y + g \cdot \mathbf{1}[a_i = D]$, net of a fixed policy loss.

Assumption 1 (A cooperation problem)**.** $\phi\kappa < g < 2\phi\kappa$.

The left inequality makes deviation individually tempting: the private policy gain g exceeds a faction's own share $\phi\kappa$ of the support it costs. The right makes unity jointly optimal: the total support hit $2\phi\kappa$ from a deviation exceeds its gain. So (A, A) maximises joint payoffs but is not a stage-game equilibrium; the externality – the deviator ignores the $\phi\kappa$ it imposes on its partner – is what must be disciplined over time. Indeed, since $g > \phi\kappa$, deviation strictly dominates in the stage game – the net gain $g - \phi\kappa$ is collected whatever the partner does – so (D, D) is the unique stage-game Nash equilibrium. This is the state to which the split phase reverts; the punishment is a return to stage-game play, credible without any off-path threat, exactly as Cournot reversion sustains the cartel in Green and Porter (1984). Within a period the two factions choose simultaneously; support y is then realised and publicly observed, and play moves on. Histories record past support, never past actions: a faction never sees whether its partner held the line, only the support that resulted. A low y is for that reason ambiguous – a defection ($d \geq 1$), or merely an adverse draw of ε – and it is this ambiguity the discipline must work around.

3 Equilibrium: discipline by the threat of a split

We study the trigger-strategy equilibrium in the manner of Green and Porter (1984). Play opens in a *unity* phase, (A, A), and remains there while support clears a trigger, $y \geq \hat{y}$. The first time $y < \hat{y}$, play switches to a *split* phase – both factions advocate their own lines, (D, D), the party visibly divided at depressed support $\bar{s} - 2\kappa + \varepsilon$ – for T periods, after which unity resumes; Figure 1 depicts a sample path.

Let V_C and V_P be a faction's discounted payoffs at the start of the unity and split phases. On the path both advocate, so $d = 0$ and a split is triggered with probability

$$q_0 = \Pr[\,\bar{s} + \varepsilon < \hat{y}\,] = \Phi\big((\hat{y} - \bar{s})/\sigma\big) > 0. \tag{2}$$

Consider a faction weighing a one-shot deviation in a unity period – deviating now, reverting to advocacy thereafter. It collects the net gain $g - \phi\kappa$ today (the policy gain, less its own share of the support it costs); but by lowering expected support by κ it raises the probability of tripping the trigger from q_0 to $q_1 = \Phi\big((\hat{y} - \bar{s} + \kappa)/\sigma\big) > q_0$. Unity is incentive-compatible exactly when

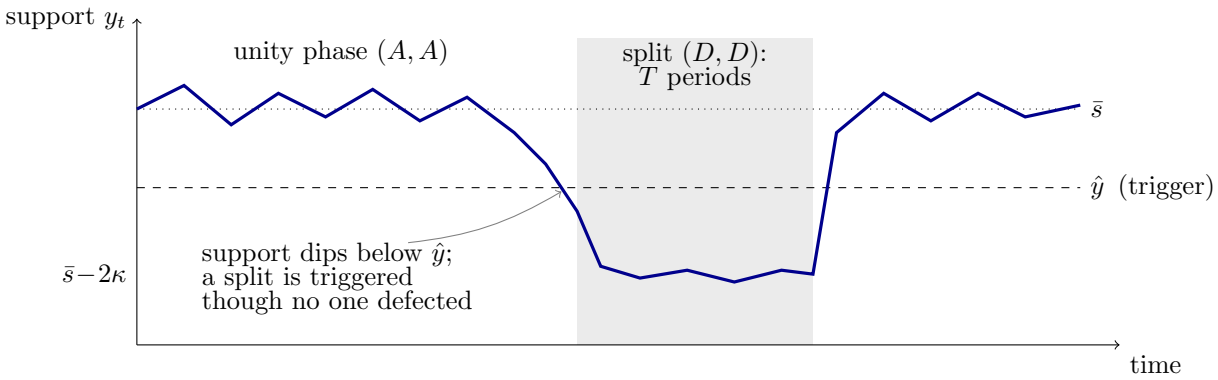

Figure 1: The model and the trigger strategy. The factions advocate the common line while the party's support $y_t = \bar{s} - \kappa d + \varepsilon$ clears the trigger $\hat{y}$. A run of adverse draws can carry support below $\hat{y}$ even when both advocate ($d = 0$); the party then splits – each faction advocating its own line, support falling to $\bar{s} - 2\kappa$ – for T periods, after which unity resumes. The split on the equilibrium path falls on factions that have not defected.

that one-shot gain does not exceed the discounted cost of the extra split-risk it creates:

$$\underbrace{g - \phi\kappa}_{\text{temptation}} \;\leq\; \delta\,(q_1 - q_0)\,(V_C - V_P). \tag{3}$$

Proposition 1 (No discipline without splits)**.** *Whenever unity is sustained by* (3)*, splits occur on the equilibrium path with probability $q_0 > 0$ each period, triggered by support rather than by any observed defection. There is no equilibrium of this class that holds the factions to the line and never splits: setting $q_0 = 0$ (never punishing) makes the right-hand side of* (3) *vanish while the temptation $g - \phi\kappa > 0$ remains, so each faction deviates and unity unravels. The split a party suffers when no faction has betrayed it is the price of the discipline that keeps them loyal.*

4 The optimal scheme

There is no planner here, only two factions; the scheme is whichever trigger-strategy equilibrium they play. We select the one they would agree to behind the symmetry of their situation – the Pareto-efficient symmetric public-perfect equilibrium – in the manner of the best collusive equilibrium in the cartel. With the factions ex ante alike, this is the natural focal scheme, and the comparative statics below are its.

Write w_C, w_P for the per-period payoffs in the unity and split phases, with $w_C > w_P$: a split is costly even though each faction, advocating its own line, enjoys a moment of policy relief, because the electoral loss outweighs it. The phase values satisfy

$$V_C = w_C + \delta\big[q_0 V_P + (1 - q_0)V_C\big], \qquad V_P = w_P\,\frac{1 - \delta^T}{1 - \delta} + \delta^T V_C. \tag{4}$$

Substituting the second into the first and rearranging, these yield two transparent expressions. The deterrent – the value a faction forfeits by tipping the party into a split – is

$$\Delta \equiv V_C - V_P = \left(1 - \delta^T\right)\left(V_C - \frac{w_P}{1-\delta}\right), \tag{5}$$

rising in the length T (a longer split deters more) and bounded above by the grim value at $T = \infty$. We posit a reversion of fixed length, but it is the deterrent Δ that matters; its largest credible value via reversion to stage-game play is this grim bound $\Delta_{\max}$, attained as $T \to \infty$. (Harsher self-generating punishments in the sense of Abreu, Pearce, and Stacchetti (1990) exist in principle, but reversion to the unique stage-Nash (D, D) is credible without any off-path threat, and is what we use.) And the annuitised value of unity is

$$(1-\delta)V_C = w_C - \delta\, q_0\, \Delta, \tag{6}$$

its first-best w_C less the per-period cost of discipline $\delta q_0 \Delta$ – the splits the party suffers, at rate q_0, when no faction has defected. To maximise welfare is to minimise that deadweight $q_0\Delta$ subject to incentive compatibility, which we now write $\delta\, D(\hat{y})\, \Delta \geq g - \phi\kappa$ with detection power $D(\hat{y}) \equiv q_1 - q_0$.

Two facts shape the solution. The false-alarm rate $q_0 = \Phi\big((\hat{y} - \bar{s})/\sigma\big)$ rises with the trigger. Detection power $D(\hat{y}) = \Phi\big((\hat{y} - \bar{s} + \kappa)/\sigma\big) - \Phi\big((\hat{y} - \bar{s})/\sigma\big)$ is single-peaked in it – nil for a trigger never or always tripped, greatest in between. Because a costly split should never be harsher than deterrence requires, the constraint binds, fixing $\Delta = (g - \phi\kappa)/\delta D(\hat{y})$; substituting into the deadweight, the party's problem collapses to

$$\min_{\hat{y}} \ \frac{q_0(\hat{y})}{D(\hat{y})}. \tag{7}$$

Proposition 2 (The optimal split is rare and hard)**.** *The optimal trigger minimises the ratio of false alarms to detection, q_0/D. It is lenient: the party forgives as much as deterrence allows, holding the frequency of false-alarm splits down and compensating with their length – so the optimal split is rare and severe, not frequent and mild.*

The severity at the corner is an artefact of the symmetric punishment, and it is what the renegotiation problem of §6 acts on; an asymmetric, renegotiation-proof split would temper it. The comparative statics are already legible in (7) and the binding constraint.

5 Comparative statics

The signs follow from the binding constraint $\delta D(\hat{y})\Delta = g - \phi\kappa$ together with (7).

Proposition 3 (When parties split)**.** *The optimal scheme splits more often, or for longer, as* (i) *the factions' ideals diverge – g rises with $|x_R - x_L|$, tightening* (3)*;* (ii) *the support signal grows noisier – larger σ shrinks $q_1 - q_0$ for given $\hat{y}$, so the trigger must be raised; and splits less often as* (iii) *the factions grow more patient (δ) or office more valuable (ϕ), which slacken* (3)*.*

The first is the substantive prediction: divided parties split more, not because division mechanically fractures them but because wider ideals raise the temptation the discipline must offset. The second is that splits are partly an artefact of noise. Parties whose electoral fortunes are hard to read punish themselves more for slumps they cannot attribute.

This last comparative static has an institutional reading. The office value ϕ is, concretely, the rate at which the spoils of size accrue to a party in proportion to its support – public funding above all, in the manner of the cartel party of Katz and Mair (1995), but also parliamentary group status, committee seats, and access to the airwaves. The cost of a split is then the subvention a divided or shrunken party forfeits, and the comparative static says that the more generously, and the more steeply, a regime rewards size, the fewer the splits, so that size-proportional public finance supports party discipline. Where the schedule is convex, or carries a threshold, such as a minimum size for group status and its grants, the effect is larger, since a split then destroys value rather than merely dividing it, and a party near such a threshold is held together by it. This is a second route, alongside factional divergence, to the survival threshold of §9. The prediction is cross-national: party systems with generous, size-sensitive public funding should have more disciplined, less fissile parties, and reforms that flatten the schedule should loosen them.

6 Renegotiation

The split of Section 3 is strongly symmetric: it hurts both factions, and both would prefer to skip it and reunite at once. It is therefore Pareto-dominated by its own continuation and fails the renegotiation-proofness of Farrell and Maskin (1989). This is not a technical blemish but a substantive point: a leader able to broker immediate reunification, dismissing the split as "surely just a bad poll," destroys the threat that held the factions in line, and discipline collapses. A party that would stay united must make its splits *non-renegotiable*: the credible prospect of a division nobody wanted is what prevents the divisions the factions would otherwise choose. The discipline rests on commitment to endure the costly outcome.

7 Renegotiation-proof splits: factional capture

The symmetric split of §6 fails because it is symmetric: hurting both, it invites them to reunite. Consider an alternative. When support falls below the trigger, let the line pass not to division but to one faction, chosen, since the signal names no culprit, by a public coin, which sets the

party's line for the punishment phase while the other is sidelined. Call this *factional capture*. It is renegotiation-proof: the favoured faction strictly prefers its turn in charge to the compromise, so it will not consent to reunify, the continuation is not Pareto-dominated, and the punishment is carried out because one faction wants it carried out, following the logic of Farrell and Maskin (1989) and of Abreu's asymmetric penal codes.

Why the lottery escapes the renegotiation that undid the symmetric split is worth setting out. The coin is drawn by a standing rule of succession: who holds the line in a crisis is fixed in advance, not bargained over once the crisis comes. It therefore falls before the factions can re-bargain, and it produces a winner with a stake in seeing the punishment through. The split is then a lottery over the line: when support drops, the leadership passes to one wing for a time. That wing prefers its turn and will not surrender it, so no reunification carries both signatures; and the sidelined wing accepts its season because the next draw may fall its way. The prospect of holding the line next time is what sustains the discipline. The symmetric split offered no such prospect, only a shared cost, and so could always be renegotiated. The lottery, by giving each faction a standing chance at the line, gives each a reason to keep the punishment in place.

Deterrence survives the failure to identify a culprit. A faction that deviates raises the chance of the capture phase, in which the coin leaves it sidelined with probability one-half; the prospect of losing the party to its rival deters, though no one is ever convicted. Because the punishment is desirable to its beneficiary, it need not be grim: a finite, moderate capture suffices, which removes the corner of Proposition 2.

The phase, specified. Concretely, when a capture for f is triggered, for T periods the favoured wing f advocates its own ideal (plays D) while the sidelined wing s holds the manifesto (plays A); support falls by a single κ ($d = 1$), not the 2κ of mutual paralysis. Per period the favoured wing earns $w_f = w_C + (g - \phi\kappa) > w_C$ – it is the unpunished deviator – and the sidelined wing earns $w_s = w_C - \phi\kappa < w_C$, bearing the electoral hit while it holds the line. So entering a capture is worth v^f to the favoured wing and v^s to the sidelined, with $v^f > V_C > v^s$: a shift to one wing, electorally milder than a symmetric split and policy-favourable to its beneficiary. The favoured wing plays its dominant action and needs no inducement; the sidelined wing is held to A by the prospect that grabbing policy now only lengthens its own sidelining – an extension the favoured wing is glad to enforce, so discipline within the phase is itself renegotiation-proof.

Proposition 4 (Capture, not paralysis)**.** *An asymmetric punishment that hands the line to a randomly chosen faction is weakly renegotiation-proof and sustains unity; under it splits take the form of temporary factional capture rather than mutual paralysis, and the optimal punishment is finite.*

This is the more faithful reading of politics. A party in trouble does not usually freeze in two; it shifts to one wing. That shift is what makes the discipline credible, because the wing it shifts to prefers it.

The optimal lottery. The coin need not be fair, and the model identifies which bias best disciplines. Let the line pass to L with probability p and to R with $1-p$; a larger p leaves L favoured more often – milder for L, harsher for R – so it shrinks L's deterrent and swells R's.

Proposition 5 (The optimal lottery)**.** *Let the capture award the line to L with probability p. Write $A = V_C - v^s$ and $B = v^f - v^s$ for the unity value net of the sidelined and the favoured capture values, with $0 < A < B$, and $\theta_i = (g_i - \phi\kappa)/[\delta(q_1 - q_0)]$ for each faction's required deterrent. Unity is sustainable for some p iff $1 - (A - \theta_R)/B \le (A - \theta_L)/B$; when it is,* (i) *with equal temptations the fair coin $p = \frac{1}{2}$ is optimal, maximising the smaller of the two deterrents;* (ii) *with unequal temptations the discipline-optimal lottery is biased against the more-tempted faction; and* (iii) *no sustaining bias exceeds $A/B < 1$ in either direction – a coin that nearly always crowns one wing ceases to deter it, the favoured wing preferring the capture it expects to win.*

The fair coin thus sits at the centre of a range of workable lotteries, with discipline bearing most on the wing most inclined to defect; and a party whose rule of succession tilts to a dominant wing is, by the same logic, the party least able to discipline that wing.

8 Correlated fortunes and the attribution of blame

We have let the party see only its aggregate support, so that a slump can be pinned on neither faction. Suppose instead it sees each wing's electoral contribution,

$$y_i = \bar{s}_i - \kappa\,\mathbf{1}[a_i = D] + \varepsilon_i, \qquad \varepsilon_i = u + \eta_i, \quad i \in \{L, R\},$$

where u is a shock common to both wings – the party's shared fortune – and η_i is a wing's own. The correlation of the two fortunes, $\rho = \mathrm{Var}(u)/\mathrm{Var}(\varepsilon_i)$, measures how much of its fate a wing shares with its partner.

The aggregate of Section 2 throws attribution away; the *difference* recovers it. Since

$$y_L - y_R = (\bar{s}_L - \bar{s}_R) - \kappa\big(\mathbf{1}[a_L = D] - \mathbf{1}[a_R = D]\big) + (\eta_L - \eta_R),$$

the common shock u cancels, and a unilateral deviation – one wing free-riding while the other holds the line, the very deviation the discipline must deter – shifts the difference by κ against noise of variance $2\,\mathrm{Var}(\eta_i)$. As the fortunes grow more correlated ($\rho \to 1$, the idiosyncratic part vanishing), that noise goes to zero and the difference becomes a clean signal of which wing defected.

The signal does more than diagnose; it directs the punishment. Replace the coin of §7 by the difference. On a punishment trigger, form the demeaned difference $\hat{\delta} = (y_L - y_R) - (\bar{s}_L - \bar{s}_R)$ and hand the line, for the phase, to the wing the difference exonerates – to R when $\hat{\delta}$ falls low

enough to mark L, to L when it runs high enough to mark R – sidelining the wing it accuses; where $\hat{\delta}$ accuses no one, fall back on the coin. This is a *targeted capture*: a capture of the line exactly as in §7, and renegotiation-proof for the same reason, since the exonerated wing prefers to govern and will not reunify, but aimed by the signal rather than by chance. A faction is now deterred by the prospect of being named and sidelined rather than by that of a collective split: a deviation shifts $\hat{\delta}$ by κ toward its own accusation and so raises the chance the line is taken from it.

Proposition 6 (Attribution and targeted capture)**.** *Let the party punish a low signal by the targeted capture above. It is weakly renegotiation-proof, and it sustains unity whenever*

$$g - \phi\kappa \;\le\; \delta\,(\pi_1 - \pi_0)\,\big(V_C - V_P^s\big),$$

where $\pi_1 - \pi_0 > 0$ is the rise in the probability a deviator is sidelined and $V_C - V_P^s$ the value it forfeits when sidelined; $\pi_1 - \pi_0$ is increasing in κ/σ_η, and so in ρ. As $\rho \to 1$ the deviator is named with certainty, the capture falls on the guilty wing alone, and the blameless on-path split vanishes; as $\rho \to 0$ the difference is uninformative and the scheme reduces to the collective trigger of Section 3. The false-alarm split is the mark of factions whose fortunes are independent.

This supplies the evidence the capture of §7 lacked. In that case a slump named no culprit and the punishment fell by a coin; with a shared fortune the party can difference it out and direct the punishment accurately. It also adds a second axis of factional difference. Wings may differ in their ideals, which raises the temptation g and strains discipline; or in their electoral base, whose independence, a low ρ, deprives the monitor of attribution and forces the collective split. A party of wings that think alike but draw on separate constituencies splits blamelessly; a party of wings that think differently but court the same voters can identify the guilty and discipline them. Diversity of mind and diversity of base affect party discipline in opposite directions.

9 The endogenous manifesto

We have held the line m, and so the temptation g, fixed. Let it be chosen. A faction's gain from deviating grows with the distance of the line from its ideal: write $g_i = g(|m - x_i|)$, increasing, so a line far from a faction tempts it more. The binding constraint is that of the more tempted faction, $\max_i g(|m - x_i|)$, and the split rate rises with it. Two things follow.

First, the manifesto is pulled to the centre by discipline, not only by the electorate. The line that holds the party together most easily equalises the temptations – the midpoint $m^* = \frac{1}{2}(x_L + x_R)$ when g is symmetric – since moving toward either faction slackens its constraint only by tightening the other's binding one. A party whose manifesto is captured by one wing makes the other wing hard to hold, and splits the more for it.

Second, unity can fail outright. Even the best line leaves a temptation $g\big(\frac{1}{2}|x_R - x_L|\big)$, so

there is a critical divergence $\bar{D}$ beyond which no manifesto satisfies the incentive constraint: a party whose factions stand too far apart cannot be disciplined by any line, and the schism is permanent. Below $\bar{D}$ the party holds and splits occasionally; above it, the party divides for good.

Proposition 7 (The centripetal manifesto and the survival threshold)**.** *The discipline-optimal manifesto minimises the maximum temptation, drawing the line toward the centre of the party; and there is a critical factional divergence $\bar{D}$ above which no manifesto sustains unity, so that sufficiently divided parties split for good.*

10 Empirical implications

The model speaks to data on when, how often, and in what form parties divide. Its readings differ from the conventional ones, which treat a split as the surfacing of policy disagreement or as a failure of leadership, and they can be stated as hypotheses.

H1 (Trigger). Because the party punishes the signal and not the act, splits follow electoral downturns rather than detectable betrayals: a poll slump, a lost by-election, or a shock with no factional author, such as a recession or an unrelated scandal, can set off a split while every faction has held the line. Division is often a response to bad luck rather than bad faith. *H2 (Noise).* Parties whose electoral fortunes are harder to read split more: the noisier the link between unity and the vote, the more a party punishes slumps it cannot attribute, and the more it splits without cause – even holding the wings' divergence fixed. *H3 (Divergence).* Divided parties split more, but through discipline, not mechanics: a wider gap between the wings raises the temptation the discipline must offset, and with it the rate of splits. *H4 (Form).* A disciplined party splits rarely and hard rather than often and mildly, and resolves a crisis by capture – the line passing to one wing for a time – not by symmetric paralysis. *H5 (Threshold).* Below a critical divergence the party survives, moderating its manifesto to hold together; above it no manifesto disciplines the factions and the schism is permanent. The threshold widens with the value of office and the patience of the factions and narrows with electoral noise, so a governing party can hold a wider range of factions than a hopeless one.

Measurement, and a role for text. The model's objects are observable, several from text. The agreed line is the party's manifesto or the leadership's stated position; a faction's advocacy of it, as against its drift toward its own ideal, is the distance between the faction's own communication, the parliamentary speech of its members, its motions, and its press, and that line, scaled by the now-standard text-as-data methods (Laver, Benoit, and Garry, 2003; Slapin and Proksch, 2008; Grimmer and Stewart, 2013). The divergence of H3 is then the dispersion of the wings' positions, measured by a rising within-party dispersion of legislators' speech; the electoral noise of H2 is the volatility of the party's vote or polling; and a split is the observed

breakaway, defection, or mass switch.

The data. Each has a source. Within-party position and its dispersion come from manifestos (Volkens et al., 2021) or from large parliamentary-speech corpora such as ParlSpeech (Rauh and Schwalbach, 2020), in which the divergence of a party's wings can be followed across a parliament; electoral noise from the volatility of vote and poll series; and splits, defections, and legislative switches from the established records of party change (Laver and Benoit, 2003; Desposato, 2006; Ceron, 2014).

An estimation, and what would distinguish the model. The pieces compose into a hazard of split, its covariates the within-party dispersion of positions (H3), the volatility of the party's electoral signal (H2), and a recent poll slump or adverse shock (H1), with the threshold of H5 entering as the sharp nonlinearity by which the hazard climbs once dispersion passes a critical width. What distinguishes the account from the literature that reads division as deliberate exit (Ceron, 2014; Laver and Benoit, 2003; Desposato, 2006; Izzo, 2024) is the *blameless* split: the model predicts splits with no antecedent rise in preference divergence, splits on electoral noise alone, and predicts that noisier parties split more even where the wings have not moved apart. That literature explains the division a disagreement produces; the test looks for the division a bad draw produces, on factions that had not moved.

A test, and what would settle it. We leave the estimation to a companion empirical paper and sketch only the design. The natural object is a discrete-time hazard of a split on a panel of party-spells – a complementary log-log with duration polynomials and party-clustered errors – whose covariates are the within-party dispersion of positions (H3), the volatility of the party's electoral signal (H2), and a recent slump or exogenous shock (H1), with the threshold of H5 entering as a sharp nonlinearity by which the hazard climbs once dispersion passes a critical width. The distinctive test is for the blameless split: an elevated split hazard with no antecedent rise in preference divergence, and a higher hazard for noisier parties holding divergence fixed. The model expects the unreadability of the signal, rather than its level, to drive the split, and this prediction, division on a noise the party cannot tell from disloyalty, separates the account from theories of division as deliberate exit.

Two predictions are the most direct to bring to data. The within-party divergence of H3 calls for the dispersion of the wings' positions across a parliament; and the threshold of H5, whether a split is the temporary kind a party recovers from or the permanent schism beyond $\bar{D}$, calls for splits coded by type, a cause-specific hazard. The claim that organises the design is the distinctive one: the split falls on the party whose fortunes are hardest to read.

11 Three schisms

The mechanism is also a reading of particular splits. Three cases illustrate its different implications.

Labour and the SDP, 1981 – the survival threshold. Through 1979–81 the distance between Labour's social-democratic right and its ascendant left widened on issue after issue – unilateral disarmament, withdrawal from the European Community, the Wembley electoral college – until no manifesto could be found that held both wings within reach of the discipline. The Gang of Four left and formed the Social Democratic Party, which never returned. This is divergence past the critical $\bar{D}$ of Proposition 7: not a split the party recovers from, but the permanent schism that follows when the factions stand too far apart for any line to hold them.

The Liberals and Home Rule, 1886 – the line no centre could hold. A party may be broken by a single line it cannot place, rather than by slow drift. Gladstone's Home Rule Bill set the party a question on which neither the Whigs under Hartington nor the radicals under Chamberlain could be held to the leadership; ninety-three Liberals crossed, and the Liberal Unionists broke away to the Conservatives. In the model's terms the endogenous manifesto found no interior point that kept the most-tempted wing within its constraint, since the issue fixed a divergence beyond what discipline could offset, and a party long divided between Whig, radical, and Gladstonian came apart on it.

The SPD, 1914–1917 – the shock with no author. One of the model's central claims is that a split can fall on a party none of whose factions has betrayed it, triggered by an event outside it. The war credits of August 1914 set the German Social Democrats a line – support for the war, the *Burgfrieden* – that no faction had chosen and that opened a rift between a pro-war majority and an anti-war minority. The truce held for a time, as the discipline holds through a single bad draw; but the pressure did not relent, and in 1917 the minority left to form the Independent Social Democrats. The war was the noise the party could not tell from disloyalty, and the schism that followed was its consequence.

12 Related literature

The paper draws on four literatures: the economics of collusion, which supplies its mechanism; the theory of parties as cartels, whose central claim it formalises; the positive study of party splits, whose explananda it recasts; and work on cohesion, factions, and leadership, to which it connects most directly.

Collusion under imperfect monitoring. The mechanism is that of Green and Porter (1984): a cartel that observes only a noisy public signal cannot tell a slump from cheating, and so triggers price wars on the path, punishing slumps it knows may be innocent because to forgive every doubtful one is to invite the cheating it cannot see (the empirical reading is Porter,

1983; the self-generating bound on such punishments is Abreu, Pearce, and Stacchetti, 1990; that collusion is hardest to hold when the joint enterprise is most strained is the countercyclical theme of Rotemberg and Saloner, 1986, the natural comparison for splits that follow downturns). The closest case by structure is *collusion among equals*, where the punishment is mutual reversion rather than a principal's sanction: Bagwell and Staiger (1990) sustain tariff cooperation between two governments by the threat of an on-path trade war set off by noisy trade volumes, as our factions punish one another with a split. A related but distinct line carries the Green–Porter signature, a punishment that falls on the path on the diligent, into principal–agent settings: Padró i Miquel and Yared (2012), where a sovereign disciplines a client it monitors only through the disturbances it sees, and Acharya, Lipnowski, and Ramos (2025), where voters discipline a politician whose effort they infer from noisy outcomes. These are a principal's discipline of an agent, not collusion among equals; ours is the latter, and the split is mutual. We bring the Green–Porter logic to a place it has not been taken, the internal discipline of the party, and add three features these settings have no room for: a line the cooperating factions themselves choose, so that the manifesto is pulled to the centre and a wide enough divergence breaks the party outright (§9); a renegotiation-proof punishment that takes the political form of capture rather than mutual reversion (§7); and the attribution of blame when factional fortunes move together (§8).

Parties as cartels. The motivating claim is that significant party behaviour is organised behaviour – the majority party acting as a legislative cartel, monopolising the agenda to protect a collective good (Cox and McCubbins, 1993, 2005) – with its European counterpart in the cartel party of Katz and Mair (1995), where the good is the public subvention that size commands. The sceptic's challenge is Krehbiel (1993): much that looks like party influence is only the shared preferences of members who would have voted alike in any case. The cartel thesis answers the challenge but has not been given the theory its name invokes. It has not been set down as a repeated game, and it offers no account of the breakdowns observed on the path. This paper supplies both: the cartel as a Green–Porter game of imperfect monitoring, with the on-path split as the phenomenon the thesis could not accommodate. The account sits alongside other microfoundations of the party, as a solution to politicians' collective dilemmas (Aldrich, 1995), a shared brand (Snyder and Ting, 2002), a commitment device that widens what members can promise (Levy, 2004), a body shaped by the electoral rule (Morelli, 2004) and by the internal distribution of power (Invernizzi and Prato, 2025). Where these explain why parties cohere, we explain the discipline they must run, and the splits it costs, to do so.

Party splits, dissent, and switching. The positive literature on division models it as a deliberate act: the exit of factions or legislators who prefer to go, for policy (Ceron, 2014) or for office and ambition (Laver and Benoit, 2003; Desposato, 2006); the electorally costly dissent a faction wields to sway a learning electorate (Izzo, 2024); the evolution of parties as

members' incentives shift (Invernizzi and Izzo, 2024). Our split differs from these. It falls, on the equilibrium path, on factions that would rather stay and that have chosen nothing; the discipline falls on the loyal. That literature explains the division a disagreement or a strategy produces; we explain the division that discipline requires. The two are complements, the observed splits a mixture of the deliberate and the disciplinary.

Cohesion, factions, and leadership. The paper connects most closely to a programme on how parties and governments hold together. Cohesion enforced by procedure – the vote of confidence – is the subject of Diermeier and Feddersen (1998); here the discipline runs not through a constitutional device but through the factions' own trigger strategy. Dewan and Squintani (2016) reads factions not as pathologies but as bearers of information, an informational reading of parties more broadly (Snyder and Ting, 2002), and Persico, Rodríguez-Pueblita, and Silverman (2011) reads them as the organised competitors for resources within a party; we take them as the disciplining units themselves, each holding the other to the line. Leadership enters here too. Dewan and Myatt (2007) make the leader a coordinator of the party's direction, and our renegotiation problem (§6) gives that role a further implication: a leader able to broker immediate reunification, to dismiss a split as "just a bad poll," destroys the discipline, so that the leadership which sustains a party in one model can dissolve it in this one. The account also has a companion: the extension to three or more factions, where capture and attribution become the formation and survival of coalitions, is developed separately.

13 Concluding remarks

We have modelled a party as two factions that sustain a common manifesto only under imperfect monitoring. Discipline of this kind cannot do without the occasional split. The party punishes a fall in support it cannot tell from a defection, so splits fall on the blameless, and these splits are what keep the factions loyal. The optimal split is rare and severe. Made robust to renegotiation it becomes a temporary capture of the line by one wing rather than a mutual paralysis. And the manifesto that best holds the party together is drawn to its centre, with a divergence beyond which no line holds and the party breaks for good.

We have kept the model deliberately spare: two factions, a symmetric stage game. The natural next step is to go beyond two, to ask how capture and attribution work among several factions, where coalitions form and the difference of fortunes no longer names a single culprit. The central point is already clear. A party holds together not by never dividing but by managing how it divides. The credible prospect of a split that nobody wants is what forestalls the splits that the factions would otherwise choose.

References

Abreu, D., D. Pearce, and E. Stacchetti (1990). Toward a Theory of Discounted Repeated Games with Imperfect Monitoring. *Econometrica* 58(5), 1041–1063.

Acharya, A., E. Lipnowski, and J. Ramos (2025). Political Accountability under Moral Hazard. *American Journal of Political Science* (forthcoming).

Aldrich, J. H. (1995). *Why Parties? The Origin and Transformation of Political Parties in America*. University of Chicago Press, Chicago.

Bagwell, K., and R. W. Staiger (1990). A Theory of Managed Trade. *American Economic Review* 80(4), 779–795.

Ceron, A. (2014). Gamson Rule Not for All: Patterns of Portfolio Allocation among Italian Party Factions. *European Journal of Political Research* 53(1), 180–199.

Cox, G. W., and M. D. McCubbins (1993). *Legislative Leviathan: Party Government in the House*. University of California Press, Berkeley.

Cox, G. W., and M. D. McCubbins (2005). *Setting the Agenda: Responsible Party Government in the U.S. House of Representatives*. Cambridge University Press, Cambridge.

Desposato, S. W. (2006). Parties for Rent? Ambition, Ideology, and Party Switching in Brazil's Chamber of Deputies. *American Journal of Political Science* 50(1), 62–80.

Dewan, T. (2002). The Party's Over. Doctoral thesis, Nuffield College, University of Oxford.

Dewan, T., and D. P. Myatt (2007). Leading the Party: Coordination, Direction, and Communication. *American Political Science Review* 101(4), 827–845.

Dewan, T., and F. Squintani (2016). In Defence of Factions. *American Journal of Political Science* 60(4), 860–881.

Diermeier, D., and T. J. Feddersen (1998). Cohesion in Legislatures and the Vote of Confidence Procedure. *American Political Science Review* 92(3), 611–621.

Farrell, J., and E. Maskin (1989). Renegotiation in Repeated Games. *Games and Economic Behavior* 1(4), 327–360.

Green, E. J., and R. H. Porter (1984). Noncooperative Collusion under Imperfect Price Information. *Econometrica* 52(1), 87–100.

Grimmer, J., and B. M. Stewart (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis* 21(3), 267–297.

Invernizzi, G. M., and F. Izzo (2024). Evolving Parties. Working paper.

Invernizzi, G. M., and C. Prato (2025). Bending the Iron Law. *American Journal of Political Science* (forthcoming).

Izzo, F. (2024). With Friends Like These, Who Needs Enemies? A Model of Electorally Costly Dissent. *Journal of Politics* 86(3).

Katz, R. S., and P. Mair (1995). Changing Models of Party Organization and Party Democracy: The Emergence of the Cartel Party. *Party Politics* 1(1), 5–28.

Krehbiel, K. (1993). Where's the Party? *British Journal of Political Science* 23(2), 235–266.

Laver, M., and K. Benoit (2003). The Evolution of Party Systems between Elections. *American Journal of Political Science* 47(2), 215–233.

Laver, M., K. Benoit, and J. Garry (2003). Extracting Policy Positions from Political Texts Using Words as Data. *American Political Science Review* 97(2), 311–331.

Levy, G. (2004). A Model of Political Parties. *Journal of Economic Theory* 115(2), 250–277.

Morelli, M. (2004). Party Formation and Policy Outcomes under Different Electoral Systems. *Review of Economic Studies* 71(3), 829–853.

Padró i Miquel, G., and P. Yared (2012). The Political Economy of Indirect Control. *Quarterly Journal of Economics* 127(2), 947–1015.

Persico, N., J. C. Rodríguez-Pueblita, and D. Silverman (2011). Factions and Political Competition. *Journal of Political Economy* 119(2), 242–288.

Porter, R. H. (1983). A Study of Cartel Stability: The Joint Executive Committee, 1880–1886. *Bell Journal of Economics* 14(2), 301–314.

Rauh, C., and J. Schwalbach (2020). The ParlSpeech V2 Data Set: Full-Text Corpora of 6.3 Million Parliamentary Speeches. Harvard Dataverse.

Rotemberg, J. J., and G. Saloner (1986). A Supergame-Theoretic Model of Price Wars during Booms. *American Economic Review* 76(3), 390–407.

Slapin, J. B., and S.-O. Proksch (2008). A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science* 52(3), 705–722.

Snyder, J. M., and M. M. Ting (2002). An Informational Rationale for Political Parties. *American Journal of Political Science* 46(1), 90–110.

Volkens, A., T. Burst, W. Krause, P. Lehmann, N. Merz, S. Regel, B. Weßels, and L. Zehnter (2021). The Manifesto Project Dataset. Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Appendix: Proofs

Throughout write $z = (\hat{y} - \bar{s})/\sigma$ and $a = \kappa/\sigma > 0$, so that $q_0 = \Phi(z)$ and $D(\hat{y}) = \Phi(z+a) - \Phi(z)$.

Per-period payoffs. We map the stage payoff $\phi y + g\,\mathbf{1}[a_i = D]$ of Section 2 to the phase values w_C, w_P. Let ℓ_A be the (fixed) policy loss a faction bears while advocating the common line. In the unity phase both advocate, $d = 0$ so $\mathbb{E}[y] = \bar{s}$, and no deviation gain is taken, giving $w_C = \phi\bar{s} - \ell_A$. In the split phase both deviate, $d = 2$ so $\mathbb{E}[y] = \bar{s} - 2\kappa$, and each collects the deviation gain g, giving

$$w_P = \phi(\bar{s} - 2\kappa) + g - \ell_A \;\equiv\; \phi(\bar{s} - 2\kappa) - \ell_D, \qquad \ell_D = \ell_A - g < \ell_A.$$

Thus the policy loss enters both phases as the common additive term ℓ_A, the split differing only by the electoral hit $2\phi\kappa$ and the gain g. The split is net costly,

$$w_C - w_P = 2\phi\kappa - g > 0,$$

the inequality the right half of Assumption 1. Because ℓ_A is common to the two phases it cancels from every difference of values; in particular the deterrent $\Delta = V_C - V_P$, and its grim bound $\Delta_{\max} = V_C - w_P/(1-\delta)$, are independent of ℓ_A and depend on the policy primitives only through the gain g.

Lemma 1 (The detection ratio is monotone)**.** *$q_0(\hat{y})/D(\hat{y})$ is strictly increasing in $\hat{y}$.*

Proof. Write $R(z) = \Phi(z)/[\Phi(z+a) - \Phi(z)]$. Since the numerator and denominator have derivatives $\phi(z)$ and $\phi(z+a) - \phi(z)$,

$$\operatorname{sign} R'(z) = \operatorname{sign}\big(\phi(z)[\Phi(z+a) - \Phi(z)] - \Phi(z)[\phi(z+a) - \phi(z)]\big) = \operatorname{sign}\big(\phi(z)\Phi(z+a) - \Phi(z)\phi(z+a)\big),$$

the cross terms in $\Phi(z)\phi(z)$ cancelling. This is positive iff $M(z+a) > M(z)$, where $M \equiv \Phi/\phi$. Because $\phi' = -z\phi$, $M'(z) = 1 + zM(z)$; for $z \geq 0$ this is plainly positive, and for $z < 0$ the Mills inequality $|z|\Phi(z) < \phi(z)$ gives $1 + zM(z) > 0$ as well. So M is strictly increasing, $M(z+a) > M(z)$, and $R' > 0$. As $\hat{y}$ and z move together, the claim follows. □

Proof of Proposition 1. On the path both advocate, so $d = 0$ and a split occurs with probability $q_0 = \Phi(z)$. Unity is incentive-compatible only if (3) holds, whose right-hand side is $\delta(q_1 - q_0)\Delta$. A scheme that never punishes sets $\hat{y} = -\infty$, so $q_0 = q_1 = 0$ and the right-hand side vanishes,

while the temptation $g - \phi\kappa > 0$ by Assumption 1; (3) then fails and each faction deviates. Hence any scheme sustaining unity has a finite trigger and $q_0 > 0$. □

Proof of Proposition 2. By (6), maximising V_C is minimising the deadweight $q_0\Delta$. A costly split is never made harsher than deterrence requires, so (3) binds: $\Delta = (g - \phi\kappa)/[\delta D(\hat{y})]$, and $q_0\Delta = \big[(g-\phi\kappa)/\delta\big] \cdot q_0/D$. The deterrent is bounded by its grim value $\Delta_{\max} = V_C - w_P/(1-\delta)$, attained as $T \to \infty$ by (5), so feasibility requires $D(\hat{y}) \geq (g - \phi\kappa)/(\delta\Delta_{\max})$. As D is single-peaked, the feasible triggers form an interval about its peak; by Lemma 1 the objective q_0/D is increasing on it and is minimised at the interval's lenient (lower) endpoint, where the deterrence floor binds. There the trigger is lenient – q_0 small – and the deterrent is delivered by a long punishment – Δ near $\Delta_{\max}$: the split is rare and hard. Its residual severity is the artefact of the symmetric punishment lifted in Proposition 4. □

Proof of Proposition 3. At the optimum (3) binds, $\delta D(\hat{y})\Delta = g - \phi\kappa$, and the trigger minimises q_0/D over $\{\hat{y} : D(\hat{y}) \geq (g - \phi\kappa)/(\delta\Delta_{\max})\}$. *(i)* A rise in g with $|x_R - x_L|$ raises the floor $(g - \phi\kappa)/(\delta\Delta_{\max})$, shrinking the feasible interval from its lenient end; by Lemma 1 the optimal q_0 rises, and where the floor exceeds $\max_{\hat{y}} D$ the deterrent must lengthen instead – more splits, or longer. *(ii)* Since $a = \kappa/\sigma$, a larger σ lowers $D(\hat{y})$ at every $\hat{y}$, raising the same floor relative to D with the same effect. *(iii)* A larger δ raises $\Delta_{\max}$ and the weight δ; a larger ϕ lowers the temptation $g - \phi\kappa$ and raises $w_C - w_P$ and hence $\Delta_{\max}$. Each lowers the floor, widening the feasible interval toward leniency and lowering q_0 – fewer splits. □

Proof of Proposition 4. In the capture phase the favoured wing f plays D and the sidelined wing s plays A, giving per-period $w_f = w_C + (g - \phi\kappa)$ and $w_s = w_C - \phi\kappa$, so on entering the phase $v^f = \frac{1-\delta^T}{1-\delta} w_f + \delta^T V_C$ and v^s with w_s for w_f. As $w_f > w_C > w_s$, $v^f > V_C > v^s$.

Renegotiation-proofness, both directions. The continuations are unity, worth (V_C, V_C) to (f, s), and a realised capture, worth (v^f, v^s). Since $v^f > V_C$ while $v^s < V_C$, neither vector Pareto-dominates the other: renegotiating a capture back to unity is blocked by the favoured wing, which would forfeit $v^f - V_C$, and renegotiating unity into a capture is blocked by the wing that would be sidelined. No continuation Pareto-dominates another, so the scheme is weakly renegotiation-proof (Farrell and Maskin, 1989).

Within-phase incentives. The favoured wing plays its stage-dominant action D and needs no support. The sidelined wing plays A, which is stage-dominated; it is held by the rule that a within-phase deviation lengthens the capture by k periods, so that the one-shot policy gain $g - \phi\kappa$ is dominated by the discounted extra sidelining, $g - \phi\kappa \leq \delta\,\phi\kappa\,(1 - \delta^k)/(1 - \delta)$ for k large enough. The extension keeps the favoured wing in charge, which it strictly prefers, so the within-phase threat is itself renegotiation-proof and credible.

Deterrence. A one-shot deviation in a unity period raises the probability of the next capture, on entering which the fair coin sidelines the deviator with probability $\frac{1}{2}$; entering is worth

$V_P = \frac{1}{2}v^f + \frac{1}{2}v^s < V_C$, so the deterrent $V_C - V_P > 0$ enters (3) and sustains unity. Because $v^f > V_C$, the favoured wing never needs grim reversion, so a finite T suffices. □

Proof of Proposition 5. Award the line to L with probability p. Entering the capture phase, faction L is favoured (value v^f) with probability p and sidelined (value v^s) with probability $1-p$, so $V_P^L = p\,v^f + (1-p)v^s$ and its deterrent is

$$D_L \equiv V_C - V_P^L = (V_C - v^s) - p\,(v^f - v^s) = A - pB,$$

with $A = V_C - v^s$ and $B = v^f - v^s$; symmetrically $D_R = A - (1-p)B$. That $v^f > V_C > v^s$ – a faction prefers governing at its ideal to the compromise, and the compromise to being sidelined – gives $0 < A < B$. Here $B = v^f - v^s$ is the annuitised capture gain and is independent of the bias p (the $\delta^T V_C$ tails cancel), while A varies in p only through V_C, a slower channel; the comparison below is the direct effect through $-pB$. So D_L falls and D_R rises in p. The incentive constraint of faction i is $g_i - \phi\kappa \le \delta(q_1 - q_0)D_i$, i.e. $D_i \ge \theta_i$ with $\theta_i = (g_i - \phi\kappa)/[\delta(q_1 - q_0)]$. Then $D_L \ge \theta_L \iff p \le (A - \theta_L)/B$ and $D_R \ge \theta_R \iff p \ge 1 - (A - \theta_R)/B$, so a sustaining p exists iff $1 - (A - \theta_R)/B \le (A - \theta_L)/B$, the stated condition.

(i) If $\theta_L = \theta_R$ the feasible interval is symmetric about $\frac{1}{2}$, and $\min\{D_L, D_R\}$ is maximised where $D_L = D_R$, namely $p = \frac{1}{2}$. *(ii)* If L is the more tempted, $\theta_L > \theta_R$, the upper bound $(A-\theta_L)/B$ tightens while the lower bound relaxes, so the feasible set, and the slack-maximising choice within it, move to smaller p – bias against the more-tempted faction. *(iii)* For $p > A/B$ one has $D_L = A - pB < 0$, so no temptation $g_L > \phi\kappa$ satisfies L's constraint: L strictly prefers to trip the phase it expects to win. As $A < B$, $A/B < 1$, so no sustaining bias reaches the corner. □

Proof of Proposition 7. Let $g_i = g(|m - x_i|)$ with g increasing, so a line further from a faction's ideal tempts it more. The punishment is symmetric, so the deterrent $\delta(q_1 - q_0)\Delta$ is common to both factions, and (3) must hold for each; it binds on the more tempted. Unity is therefore sustainable if and only if

$$\max_i\, g(|m - x_i|) - \phi\kappa \;\le\; \delta(q_1 - q_0)\,\Delta.$$

Centripetal line. As g is increasing, $\max_i g(|m - x_i|) = g\big(\max\{|m - x_L|, |m - x_R|\}\big)$. On $m \in [x_L, x_R]$ the term $\max\{|m - x_L|, |m - x_R|\}$ equals $|m - x_L|$ for $m \ge m^*$ and $|m - x_R|$ for $m \le m^*$, a function falling then rising with a minimum at the midpoint $m^* = \frac{1}{2}(x_L + x_R)$, where both distances equal $\frac{1}{2}(x_R - x_L)$. Any line off-centre strictly raises the binding faction's temptation; the discipline-optimal manifesto is the midpoint, drawn to the centre of the party.

Survival threshold. Write $\Delta x = x_R - x_L$. The deterrent is greatest at the grim value $\Delta_{\max} = V_C - w_P/(1-\delta)$ of (5), and unity is sustainable on some line if and only if it is

sustainable on the least-tempting line m^*, that is iff

$$g\big(\tfrac{1}{2}\Delta x\big) - \phi\kappa \;\le\; \delta(q_1 - q_0)\,\Delta_{\max}.$$

By the per-period payoff accounting above, the policy loss ℓ_A is common to both phases and cancels from $\Delta_{\max}$, so divergence reaches the right-hand side by a single channel: the split-phase gain $g(\frac{1}{2}\Delta x)$, which raises w_P and hence weakly lowers $\Delta_{\max}$. The right-hand side is therefore non-increasing in Δx, while the left-hand side is continuous and strictly increasing in it (with $g(0) - \phi\kappa < 0$ at $\Delta x = 0$, where unity surely holds, and rising without bound). The two sides cross exactly once: there is a unique $\bar{D} > 0$ at which they are equal, and for $\Delta x > \bar{D}$ no manifesto – m^* included – brings the binding temptation within reach of the discipline: the party cannot hold together on any line, and the schism is permanent. □

Proof of Proposition 6. Write $\sigma_\eta^2 = \mathrm{Var}(\eta_i)$ and the demeaned difference $\hat{\delta} = (y_L - y_R) - (\bar{s}_L - \bar{s}_R) = -\kappa\big(\mathbf{1}[a_L = D] - \mathbf{1}[a_R = D]\big) + (\eta_L - \eta_R)$, free of the common shock u and distributed $N(m, \tau^2)$ with $\tau^2 = 2\sigma_\eta^2$ and mean $m = 0$ under (A, A), $m = -\kappa$ under a unilateral deviation by L, and $m = +\kappa$ under one by R.

The scheme. Fix a threshold $c \in (0, \kappa)$. On a punishment trigger, capture the line for R (sidelining L) if $\hat{\delta} < -c$, capture it for L (sidelining R) if $\hat{\delta} > c$, and revert to the coin of §7 if $|\hat{\delta}| \le c$. A capture hands the line to the exonerated wing at its ideal, so by the argument of Proposition 4 that wing strictly prefers its turn to reunification; the continuation is not Pareto-dominated and the scheme is weakly renegotiation-proof, the targeting altering who is favoured but not the credibility.

Detection. The chance L is sidelined is $\Pr[\hat{\delta} < -c] = \Phi\big((-c - m)/\tau\big)$: under (A, A) the false alarm $\pi_0 = \Phi(-c/\tau)$, under L's deviation $\pi_1 = \Phi\big((\kappa - c)/\tau\big)$. As $\kappa - c > -c$, $\pi_1 > \pi_0$, and $\pi_1 - \pi_0 = \Phi\big((\kappa - c)/\tau\big) - \Phi(-c/\tau)$ is increasing in $\kappa/\tau = \kappa/(\sqrt{2}\,\sigma_\eta)$, hence in ρ for fixed σ_u. As $\sigma_\eta \to 0$ ($\rho \to 1$), $(\kappa - c)/\tau \to +\infty$ and $-c/\tau \to -\infty$, so $\pi_1 \to 1$ and $\pi_0 \to 0$; as $\sigma_\eta \to \infty$ ($\rho \to 0$), both tend to $\Phi(0) = \frac{1}{2}$ and $\pi_1 - \pi_0 \to 0$.

Incentives. Let V_C, V_P^f, V_P^s be a faction's values of continued unity, of entering a capture favoured, and of entering it sidelined, with $V_P^f > V_C > V_P^s$. In a unity period L's one-shot deviation collects $g - \phi\kappa$ now and moves $\hat{\delta}$'s mean from 0 to $-\kappa$, raising the chance it is sidelined from π_0 to π_1 and lowering the chance it is favoured. Its continuation therefore falls by at least $\delta(\pi_1 - \pi_0)(V_C - V_P^s)$, the lost chance of being favoured only adding to the cost, so advocacy is incentive-compatible whenever

$$g - \phi\kappa \;\le\; \delta\,(\pi_1 - \pi_0)\,(V_C - V_P^s).$$

Limits. As $\rho \to 1$, $\pi_1 \to 1$ and $\pi_0 \to 0$: a deviator is named and sidelined with certainty and an innocent wing never, so the blameless on-path capture vanishes and the targeted phase carries the discipline alone. As $\rho \to 0$, $\pi_1 - \pi_0 \to 0$: the difference is uninformative, every trigger

passes through $|\hat{\delta}| \leq c$ to the coin and the aggregate trigger of Section 3, and the collective split returns. The blameless split thus marks independent fortunes. $\square$