What a Coalition Can See

A folk-theorem theory of attribution, discipline, and the survival of governments

Torun Dewan

Preliminary working note. Comments welcome.

Abstract

We develop a theory of coalitions built on the folk theorems of repeated play. Classical coalition theory is a theory of the single play. Parties form a government, divide a fixed prize, and the questions are who is in and who gets what. We treat a coalition instead as a *repeated disciplinary structure*. Parties hold a common line and a division of spoils over time, under imperfect monitoring, against the threat of collapse. Each party watches its own poll. Polls share a common national tide and carry party-specific shocks, so a slump cannot be told apart from a partner's defection. A coalition, unlike a single party, observes the whole vector of polls. Differencing the polls cancels the tide and leaves a residual that identifies a defector. The larger the coalition, the more precisely it estimates the tide it must net out. Three results follow. Attribution governs composition: a coalition can discipline only partners it can monitor, and monitoring improves with size. Coalitions are assembled for legibility rather than minimality, and may be oversized to monitor better. Disciplinability and programmability bound the viable set: a coalition governs only if its members are correlated enough to attribute and close enough to share a line. The capacity to attribute also governs survival: a coalition that uses the residual survives a national tide that would collapse a coalition relying only on the aggregate. Governments fall, on the equilibrium path, on coalitions no one chose to break. The theory comes with a programme for testing it. Its objects are measurable: the agreed line in coalition agreements, partners' positions in their manifestos, their advocacy of the line or drift from it in parliamentary text, set against the covariance of their polls. The model carries the cartel logic of Dewan (2026) from the single party to the coalition. It is set against the classical theory of formation and the opportunity theory of government termination.

1 Introduction

We develop a theory of coalitions built on the folk theorems of repeated play. A coalition is a set of parties that hold a common line and a division of spoils over time. They are disciplined as a cartel is, under imperfect monitoring and against the threat of collapse. The classical theory is a theory of the single play. A coalition forms to divide a fixed prize, offices or a policy, in one shot. Riker (1962) asks which winning coalitions are *minimal*; Baron and Ferejohn (1989), how a recognised proposer splits the spoils; Gamson (1961), why the split is proportional; Laver and Shepsle (1996), who holds which portfolio; Strøm (1990), when a party governs in the minority rather than coalesce. The questions are who is in and who gets what. Once the bargain is struck the theory says little about what follows. A government that later collapses is then explained by something the model did not contain: a duration modelled in its own right (Warwick, 1994), or a termination timed against the electoral calendar (Lupia and Strøm, 1995).

In each period a party is tempted, in the daily business of office, to break ranks and press its own line. No third party enforces the bargain. It holds, if it holds, because each member fears what its breaking would bring: the discipline of Aldrich (1995)'s party, and of the legislative cohesion that procedure enforces (Diermeier and Feddersen, 1998), carried across the parties of a government. This is the logic of a cartel under imperfect monitoring (Green and Porter, 1984). We have argued it elsewhere for the single party, whose factions discipline one another by tolerating the occasional *split* (Dewan, 2026). Here the members are parties, the split is the fall of the government, and the monitoring is concrete: each party watches its own poll.

This last feature is where the coalition is more than a larger party, and where the theory departs from the single-party case. Polls move with a *common national tide*, a mood that lifts or sinks every partner together, and with *party-specific* shocks. A tension emerges: a coalition that suffers a slump cannot tell, from its own standing alone, whether a partner has defected or the nation has turned, and yet it must, if it is to discipline anyone. It must read its partners' fortunes against its own, and it can do so. It holds not one poll but the whole vector, which carries more than the sum. Differencing two polls cancels the tide. Demeaning the whole vector estimates the tide and strips it out, and the more partners there are, the cleaner that estimate. A coalition is in this sense its own instrument of attribution, and a larger one is a better instrument.

Three results follow, and they organise the paper.

First, attribution governs who coalesces, and rewards size. A coalition can hold a partner to the common line only if it can catch that partner defecting, and it can catch a defection only if it can separate one partner's fall in support from the shared movement that lifts or lowers everyone at once. It separates the two by netting the common movement out of the vector of polls. A wider cross-section of partners estimates that common movement more precisely, so a coalition reads each of its members more clearly the more members it has. Composition becomes a selection on legibility rather than minimality. A coalition may even be *oversized*, carrying a partner it does not need for a majority, because the extra poll is not dead weight but another reading on the national mood, and so a sharper check on everyone else. This reverses the minimal-winning logic of Riker (1962) and gives the puzzle of surplus coalitions (Volden and Carrubba, 2004) an incentive rationale.

Second, disciplinability and programmability bound the set of coalitions that can govern. Two things must hold. The partners must be legible to one another: their fortunes correlated enough, or the partners numerous enough, that a defection can be told from a run of bad luck.

And they must be close enough in what they want that some common line leaves no partner so far from its own ideal that no punishment could keep it loyal. A coalition that fails the first test cannot monitor; one that fails the second cannot write a line worth holding. Either way it can be assembled, but only to come apart: outside the viable set a would-be government forms only to fall.

Third, the capacity to attribute governs survival. When a national swing drags every partner's support down together, a coalition that watches only its total vote sees the fall and cannot tell whether a partner has walked away or the whole tide has gone out; fearing the first, it breaks. A coalition that compares its partners sees their standing relative to one another unchanged, concludes that the swing was common to all, and holds. Some governments fall, then, not because dissolution became wise, but because their members could not tell that nothing had changed.

This places the paper against two literatures. Against the classical theory of coalition formation, it makes disciplinability and monitorability, rather than minimality, the principles of composition. Against the opportunity theory of government termination (Diermeier and Stevenson, 1999, 2000), in which a critical event creates the occasion to dissolve a coalition no longer worth keeping, it offers an incentive theory. The shock is not a reason to break. It is the trigger the coalition's discipline rests on, and the government it brings down may be one its members would, on reflection, have kept. The empirical content differs accordingly. Where opportunity predicts terminations the shock made sensible, we predict blameless collapses, and crises set off by a partner's relative decline rather than the coalition's absolute one.

Section 2 places the paper among the folk theorems; Section 3 sets out the model; Section 4 the discipline and the on-path collapse; Sections 5 and 6 the attribution problem, its closed form, and the targeted discipline it permits; Sections 7–9 prove the three results on composition, viability, and survival; Section 10 turns to the size of coalitions and Section 11 to coalition-proof formation; Section 12 draws out the empirics.

2 Related literature: a folk theorem for coalitions

The method is the theory of repeated games, and the results are readings of that theory's central propositions for the politics of coalition.

Cooperation as a folk theorem. That self-interested parties can hold a common line and a division of spoils, with no outside enforcer, through the shadow of what breaking would forfeit, is the folk theorem (Friedman, 1971; Fudenberg and Maskin, 1986). A governing coalition is one of its equilibria, selected, among the many the theorem permits, as the Pareto-efficient symmetric one, in the way the best collusive equilibrium is selected in a cartel.

Imperfect monitoring and the on-path collapse. A coalition does not see its members' actions, only their polls, so the relevant result is the folk theorem under *imperfect public monitoring* (Green and Porter, 1984): cooperation sustained by trigger strategies off a noisy public signal, with punishment falling *on the path*, on slumps that may be innocent, because to forgive every doubtful one is to invite the cheating the signal cannot rule out. The blameless collapse is this on-path punishment, and its severity is bounded by the worst credible self-generating punishment of Abreu, Pearce, and Stacchetti (1990), which a long enough reversion attains.

Identifiability. The closest connection is to the conditions under which imperfect monitoring still permits efficient cooperation. Fudenberg, Levine, and Maskin (1994) show that a folk theorem survives noisy public signals when the signal can statistically tell one player's deviation from another's, their *individual* and *pairwise full-rank* conditions. Our attribution gives that power political content. When partners' fortunes are correlated, the difference of their polls identifies the one that slipped, so the signal has the rank to attribute, and punishment can be *targeted*, the efficient folk-theorem outcome. When fortunes are independent the signal lacks that rank, no deviator can be named, and only the collective punishment of a collapse remains. The correlation ρ is the political measure of how near a coalition's monitoring comes to the full-rank case, and our results on composition, size, and survival turn on when that case is met: which coalitions can be policed efficiently, and which can fall back only on the collective fall of the government.

Selection: renegotiation and coalitions. Two refinements fix which equilibrium governs. A punishment both sides would cancel is not credible, and the renegotiation-proofness of Farrell and Maskin (1989) turns the symmetric collapse into the asymmetric capture or expulsion that someone is glad to carry out (Section 6). With three or more parties the threatening deviation is not only an individual's but a *sub-coalition's*, so coalition-proofness (Bernheim, Peleg, and Whinston, 1987), under which no winning subset does better apart, joins renegotiation in pinning the stable government (Section 11).

Against the coalition and survival literatures. Placed among the folk theorems, the account stands apart from the classical theory of formation (Riker, 1962; Baron and Ferejohn, 1989; Gamson, 1961; Laver and Shepsle, 1996), which is a theory of the single play. Its closest formal kin are the dynamic, stochastic bargaining models of government formation and turnover in the Baron and Ferejohn (1989) line: Merlo (1997), Diermeier and Merlo (2000), and Diermeier, Eraslan, and Merlo (2003) follow a government through time and account for its duration, its reshuffles, and its surplus partners. But their motion is re-optimisation in a changing environment the parties observe: a government turns over when bargaining makes a new deal worth more, and its terminations are efficient adaptations. Ours turns instead on imperfect monitoring: a coalition that cannot tell a slump from a defection collapses on the path, on noise. Where they derive reshuffle, surplus, and turnover from the structure of bargaining under observed shocks, we derive the same phenomena from attribution: the surplus partner kept to be seen, the reshuffle as the targeted discipline a legible coalition can manage, the collapse as the fallback of one that cannot. The same line separates us from the opportunity theory of termination (Diermeier and Stevenson, 1999, 2000), in which a shock reveals dissolution to be wise; here a shock is a trigger the discipline rests on, and the government it fells may be one its members would have kept (Section 9).

3 Model

A legislature contains parties $i = 1, \ldots, N$ with seat weights w_i and ideal points x_i on a line. A *coalition* $C \subseteq \{1, \ldots, N\}$, of size $n = |C|$, governs if it is winning, $\sum_{i \in C} w_i > \frac{1}{2}$. We take C as given through Section 9 and ask when it survives; Sections 10–11 return to its formation. The coalition has agreed a government line $m \in [\min_{i \in C} x_i, \max_{i \in C} x_i]$ and a division of the spoils of office. Time is discrete and infinite, members discounting at $\delta \in (0, 1)$.

Each period every member $i \in C$ privately chooses $a_i \in \{A, D\}$: to *advocate* the common line, or to *deviate* toward its own ideal. A deviation earns a private policy gain $g_i > 0$ but costs the coalition $\kappa > 0$ in standing. What is observed is not these choices but *polls*. Party i's poll is

$$y_i = \pi_i - \kappa \, \mathbf{1}[a_i = D] + u + \eta_i, \qquad u \sim N(0, \sigma_u^2), \quad \eta_i \sim N(0, \sigma_\eta^2) \text{ i.i.d.}, \tag{1}$$

with π_i the party's baseline standing, u a shock common to all parties – the national tide – and η_i a party-specific shock, all independent.[1] The correlation of any two partners' fortunes is

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\eta^2} \in (0, 1). \tag{2}$$

Each member draws an office return ϕY from the coalition's standing $Y = \sum_{i \in C} y_i$, $\phi > 0$, after the cartel party of Katz and Mair (1995): a government that loses support, or sheds a partner, forfeits the subvention that size commands. Office also carries a fixed prize of portfolios and patronage $\Pi > 0$, divided among the partners, Π/n each, so a larger coalition dilutes each member's share, the rivalrous spoils on which Riker (1962)'s size principle turns. A member's stage payoff is $\phi Y + \Pi/n + g_i \, \mathbf{1}[a_i = D]$, net of a fixed policy loss. The divided prize is constant within a given coalition and independent of the advocacy decision, so it cancels from the deterrent $\Delta = V_C - V_P$ and the incentive constraint (4): it leaves the discipline of Sections 4–9 untouched and enters only the choice of size, where it is the cost of breadth. We assume, as in Dewan (2026), a cooperation problem – each member is privately tempted to deviate while

[1] The tide u loads on every poll with coefficient one, which is what lets demeaning cancel it exactly (Lemma 1). With heterogeneous loadings, $y_i = \pi_i - \kappa\mathbf{1}[a_i = D] + \beta_i u + \eta_i$, demeaning leaves a residual $(\beta_i - \bar{\beta})u$, and clean attribution requires instead regressing each poll on a common factor, of which the demeaning here is the equal-loadings special case. We take equal loadings as the transparent benchmark.

joint advocacy maximises the coalition's payoff:

Assumption 1. *For every* $i \in C$, $\phi\kappa < g_i < n\,\phi\kappa$.

The left inequality makes D a dominant stage action, since the private gain exceeds the member's own share $\phi\kappa$ of the standing it costs, so universal advocacy is not a stage equilibrium; the right makes universal advocacy jointly optimal, the total cost $n\phi\kappa$ of a deviation exceeding its gain. Within a period members choose simultaneously, the poll vector $y = (y_i)_{i\in C}$ is realised and publicly observed, and play proceeds. Histories record polls, never actions. The baselines π_i are known; what a low poll cannot reveal is whether the shortfall is a defection ($-\kappa$), the tide (u), or a party's own luck (η_i).

4 Discipline and the on-path collapse

Begin with the coarse instrument, the aggregate. The coalition *cooperates*, all advocating, while $Y \geq \hat{Y}$, and the first time $Y < \hat{Y}$ the government *collapses* into a punishment phase (a caretaker interregnum, a fought election, a spell in opposition) for T periods, after which the coalition may reform. On the path all advocate, so

$$Y = \textstyle\sum_{i\in C} \pi_i + n\,u + \sum_{i\in C} \eta_i, \qquad \mathrm{Var}(Y) = n^2\sigma_u^2 + n\sigma_\eta^2. \tag{3}$$

The tide enters Y multiplied by n, since every partner moves with it, so the aggregate is, for a coalition of any size, dominated by the national tide. A collapse is triggered on the path with probability $q_0 = \Pr[Y < \hat{Y}] > 0$: the government falls when the nation turns, though no partner has defected. Let V_C, V_P be a member's discounted payoffs entering the cooperative and punishment phases, and $\Delta = V_C - V_P$ the deterrent. A member weighing a one-shot deviation collects $g_i - \phi\kappa$ now and raises the collapse probability from q_0 to $q_1 > q_0$; universal advocacy is incentive-compatible exactly when

$$g_i - \phi\kappa \;\leq\; \delta\,(q_1 - q_0)\,\Delta \qquad \text{for every } i \in C, \tag{4}$$

the binding constraint being that of the most-tempted partner, $\max_{i\in C} g_i$. The phase values, the deterrent, and the optimal trigger are as in Dewan (2026): the constraint binds, the collapse is rare and severe, and welfare is the first best less the deadweight $\delta q_0 \Delta$ of collapses suffered when no one defected. The aggregate scheme thus inherits the single party's pathology in amplified form. Because $\mathrm{Var}(Y)$ is of order $n^2\sigma_u^2$ in the tide, a coalition disciplined by its aggregate alone collapses the more readily the larger it is, and always on the national swing.

5 Attribution: reading the poll vector

The aggregate is not all the coalition sees. It holds the vector y, which carries what the sum discards: the cross-section that locates the tide and identifies the deviator. Net the known baselines and form the *residual*

$$r_i = (y_i - \pi_i) - \frac{1}{n}\sum_{j\in C}(y_j - \pi_j), \qquad i \in C. \tag{5}$$

The demeaning removes the tide entirely, whatever its size:

Lemma 1 (The tide nets out; size sharpens the residual). *Under universal advocacy, $r_i = \eta_i - \bar{\eta}$ with $\bar{\eta} = \frac{1}{n}\sum_{j\in C}\eta_j$, so the residual is free of u and distributed $N\big(0, \sigma_\eta^2(1-\frac{1}{n})\big)$. If a single member k deviates, $\mathbb{E}[r_k] = -\kappa\big(1-\frac{1}{n}\big)$ and $\mathbb{E}[r_i] = \frac{\kappa}{n}$ for $i \neq k$; the gap between the deviator's residual and any other's has mean $-\kappa$, independent of σ_u, and the per-comparison noise* $\mathrm{Var}(\eta_i - \eta_j) = 2\sigma_\eta^2$ *does not grow with n.*

Two facts about attribution sit in this lemma. First, it does not depend on the tide: the power to identify a defector depends on κ/σ_η and not on σ_u, however large the national shock. Second, it improves with size: each added partner is another draw on u, sharpening the estimate $\bar{\eta}$ of what to net out and moving the deviator's expected residual from $-\kappa(1-\frac{1}{n})$ toward $-\kappa$. Where the aggregate of Section 4 worsens with n, attribution improves with it.

A coalition that can attribute need not collapse. On a low aggregate it runs a likelihood-ratio test on the residual vector: if some member's residual is negative enough to name it the deviator beyond a chosen confidence, the coalition punishes that member alone, through expulsion, the loss of its portfolios for the punishment phase, and the line swung to the rest; if no member is named, it falls back on collective collapse. Write $p(n,\rho)$ for the probability the test correctly names a lone deviator. By Lemma 1, p is increasing in κ/σ_η – hence in ρ for fixed σ_u – and increasing in n, with $p \to 1$ as $\rho \to 1$ and $p \to$ the no-information floor as $\rho \to 0$. The deterrent in (4) is then delivered, wholly or in part, by the prospect of targeted punishment, whose expected cost to a deviator rises with p and so need not rely on the collective collapse.

6 Naming the deviator: the contrast test

The residual (5) shows attribution is possible; we now give it a closed form and turn it into a scheme that disciplines the guilty alone. The test sets each party against the rest. Define the *contrast*

$$c_i = (y_i - \pi_i) - \frac{1}{n-1}\sum_{j\in C,\ j\neq i}(y_j - \pi_j), \qquad i \in C. \tag{6}$$

Lemma 2 (The contrast). *Under universal advocacy $c_i \sim N\big(0,\, \sigma_\eta^2\frac{n}{n-1}\big)$, free of the tide; if party i alone deviates, $c_i \sim N\big(-\kappa,\, \sigma_\eta^2\frac{n}{n-1}\big)$. A deviation shifts the contrast by the full $-\kappa$, and its* *noise* $\sigma_\eta\sqrt{n/(n-1)}$ *falls with* n *toward* σ_η*. At* $n = 2$ *the contrast is the pairwise difference of Dewan (2026).*

The larger the coalition, the more precisely the other members estimate the tide that party i is judged against, and the more precise the test on i alone. Attribution sharpens with size in a single closed form, and reduces at $n = 2$ to the two-faction party.

A coalition that can attribute disciplines the guilty, not the government. Fix a threshold $c \in (0, \kappa)$. On a collapse trigger, *expel* the party whose contrast falls below $-c$, with the line and the spoils passing, for the punishment phase, to the coalition without it, and fall back on collective collapse only if no contrast does. Write

$$\pi_0 = \Phi\Big(-\tfrac{c}{\sigma_\eta}\sqrt{\tfrac{n-1}{n}}\Big), \qquad \pi_1 = \Phi\Big(\tfrac{\kappa-c}{\sigma_\eta}\sqrt{\tfrac{n-1}{n}}\Big), \tag{7}$$

for the chances a given party is named when innocent and when it alone deviated.

Proposition 1 (Targeted discipline)**.** *The expulsion scheme is weakly renegotiation-proof and sustains cooperation whenever*

$$\max_{i\in C} g_i - \phi\kappa \ \le\ \delta\,(\pi_1 - \pi_0)\,\big(V_C - V^{\text{out}}\big),$$

with V^{out} *the value of being the expelled party. The detection power* $\pi_1 - \pi_0$ *is increasing in* κ/σ_η *– hence in* ρ *– and in the coalition size* n*. As* $\rho \to 1$ *the deviator is expelled with certainty and the blameless collapse disappears; as* $\rho \to 0$ *the contrast is uninformative and discipline reverts to the collective trigger of Section 4. Because detection improves in* n*, a larger coalition is the better monitor: an added partner is a sharper test, not merely another mouth at the table.*

The composition and size results build on this scheme. It delivers the targeting that Section 7 asks of a coalition, the punishment of the guilty rather than the fall of the government, as an equilibrium, credible because the partners who remain are content to govern without the member the signal has named; and it is the closed-form $p(n, \rho) = \pi_1 - \pi_0$ that Section 10 weighs against the programming cost of breadth.

Unilateral versus joint deviation. The contrast names a lone deviator; a joint deviation, with several partners breaking ranks at once, moves their contrasts together, and the test cannot single one out. This is the right margin to police, for two reasons. The binding incentive constraint is a single party's one-shot deviation, which the contrast catches; and a coordinated break by a subset is not a defection to be disciplined within the coalition but a *breakaway* to be deterred at formation, the coalition-proofness of Section 11. What the contrast cannot attribute the aggregate still registers: a joint deviation depresses Y and trips the collective trigger of Section 4. Targeted discipline thus handles the deviation the incentive constraint turns on, with the collective collapse standing behind it for the rest.

7 Attribution and who coalesces

Proposition 2 (Attribution governs composition)**.** *A coalition can target a defecting partner – punishing the guilty rather than collapsing the government – the better it can attribute, and attribution improves with the correlation of members' fortunes and with the number of partners. As* $\rho \to 1$ *a unilateral defection is identified from the residual* (5) *and the punishment falls on the deviator alone; as* $\rho \to 0$ *no defection can be told from a partner's own bad luck, and discipline must run through collective collapse. Coalitions that must rely on collapse sustain cooperation only when the largest temptation* $\max_i g_i$ *is small; coalitions that can attribute discipline cheaply and hold more easily.*

The implication for formation follows. A party choosing partners is choosing whom it can discipline. It will coalesce where fortunes move together, through a shared electorate or an adjacent part of the spectrum, because there a partner's defection is identifiable, and it will hesitate over partners on independent trajectories, with whom every slump is ambiguous and discipline costly. Because a wider cross-section sharpens the estimate that makes a defector visible, a coalition will sometimes add a partner it does not need to win, gaining legibility with the extra poll. The principle of composition is monitorability rather than the minimality of Riker (1962); Section 10 draws out its consequence for coalition size.

8 The disciplinable programme and the survival threshold

The line m is the coalition's to choose, and with it the temptations: a member's gain from breaking ranks grows with the distance of the line from its ideal, $g_i = g(|m - x_i|)$, g increasing. The binding constraint is the most-tempted partner's, so the line that holds the coalition together most easily equalises the extreme temptations, and lies at the centre of the coalition's span.

Proposition 3 (The viable coalition)**.** *The discipline-optimal government line minimises the largest temptation, pulling the programme to the centre of the coalition; for the symmetric case it is the midpoint of the members' ideals. There is a critical spread* $\bar{D}$ *of members' ideals beyond which no line keeps the most-tempted partner's temptation within what discipline can offset, so a coalition whose members stand too far apart forms only to fall. The viable coalitions are those at once* disciplinable *–* ρ *and* n *large enough that attribution, or the collective collapse, can deter – and* programmable *– ideals spread less than* $\bar{D}$*.*

Viability is thus the intersection of two conditions that bear on different features of a coalition: monitorability turns on the correlation and number of partners, programmability on the spread of their ideals. A coalition of like-minded parties on independent electoral bases is programmable but hard to monitor; a coalition of partners who rise and fall together but sit far apart is legible but unprogrammable. The governments that endure are those that meet both conditions.

9 National shocks and the survival of governments

Proposition 4 (Surviving the tide)**.** *Whether a national shock collapses a coalition depends on its attribution capacity* (ρ, n)*. A coalition correlated and numerous enough to read the residual* (5) *meets a national swing, finds every partner's relative standing intact, infers that no one defected, and survives; its collapse hazard is governed by* κ/σ_η *and is independent of* σ_u*. As* ρ *falls the residual loses its power to attribute, and even a coalition using its best scheme is thrown back on the aggregate collective trigger of Section 4, whose hazard rises with* σ_u *and with* n*: the government then falls on the tide though every partner held the line. The blameless collapse is not the mark of a coalition that foolishly reads only the aggregate – the residual is free to any coalition with* $n \geq 2$ *– but of one whose fortunes are too independent (*ρ *low) for any scheme to tell the tide from a betrayal. Robust governments and fragile ones differ not in the shocks they meet but in whether their monitoring structure lets them see those shocks for what they are.*

Here the paper differs most clearly from the opportunity account of termination. For Diermeier and Stevenson (1999, 2000) a critical event ends a government because it reveals that dissolution has become optimal; the breakup is an improvement and a warranted one. For us a national shock ends a government because the coalition cannot tell it from a defection it must punish; the breakup is a monitoring failure, and the coalition it destroys may be one no member wished to leave and that, in hindsight, should have stood. The two accounts differ in their predictions. Opportunity predicts collapses the shock made sensible and keyed to a coalition's absolute decline. The incentive account predicts blameless collapses, and, because what a coalition punishes is a partner's standing relative to the tide it nets out, crises keyed to relative decline: a government is imperilled less by a downturn that takes all partners together than by one partner slipping against the others, even amid a general rise.

10 The size of coalitions

Composition and viability together imply a theory of coalition size that the one-shot tradition lacks. Adding a partner to C pulls three ways. It sharpens monitoring: by Lemma 1 the extra poll improves the estimate of the tide and lifts the per-comparison detection probability $p(n, \rho)$, slackening the incentive constraint (4). It strains programming: a larger coalition spans a wider range of ideals, raising the most-tempted partner's temptation $\max_i g(|m - x_i|)$. And it dilutes spoils: the fixed prize Π is split among more partners, costing each incumbent $\Pi/n - \Pi/(n+1) > 0$ – the rivalrous cost Riker (1962)'s size principle turns on, which the monitoring gain must overcome. One caution attends the first force. The expulsion test is run on all n contrasts at once, so as n grows there are more innocent partners who might be wrongly named; to hold the family-wise false-expulsion rate fixed the threshold c must tighten with n (a Bonferroni-type correction), and the net detection power, so corrected, still rises in n while σ_η is large but is bounded above: monitoring improves with size over a range, not without limit.

Proposition 5 (Optimal size, and oversized coalitions)**.** *The coalition that best sustains discipline trades the monitoring gain of an added partner against its programming and spoils-dilution costs, and is in general not minimal winning. When idiosyncratic noise σ_η is large – defections hard to read – the net monitoring gain (corrected for false expulsions) dominates the dilution and programming costs over a range, and the discipline-optimal coalition is oversized: it carries partners beyond a bare majority because the added polls are what let it attribute. When σ_η is small attribution is easy with few partners, the monitoring gain is slight, the dilution and programming costs dominate, and the optimal coalition contracts toward the minimal programmable winning set. Oversize is thus a genuine optimum against a real cost of breadth, not an artefact of costless enlargement.*

Surplus coalitions, a long-standing difficulty for minimality (Volden and Carrubba, 2004), are here a monitoring device: a government takes on a partner it does not need to win because it needs the partner to monitor. The prediction is comparative, with oversized coalitions where the electoral signal is noisy and leaner ones where it is clean, and it is the size analogue of the single party's choice of how broad a membership to keep.

The trade-off in closed form. The monitoring gain is now explicit. With the optimal symmetric threshold $c = \kappa/2$, the detection power of Section 6 is

$$p(n, \rho) = \pi_1 - \pi_0 = 2\,\Phi\!\left(\tfrac{\kappa}{2\sigma_\eta}\sqrt{\tfrac{n-1}{n}}\right) - 1, \tag{8}$$

rising in n toward its ceiling $2\Phi(\kappa/2\sigma_\eta) - 1$. A coalition of size n can discipline any temptation up to $\bar{g}(n) = \phi\kappa + \delta\, p(n, \rho)\, (V_C - V^{\text{out}})$, while the partner added to reach size n raises the temptation it must hold to $\underline{g}(n) = g\big(\tfrac{1}{2}\,\text{span}(n)\big)$, the span widening with each more distant member. The disciplinable sizes are those with $\underline{g}(n) \le \bar{g}(n)$; the optimum maximises member value within them, setting the monitoring gain and the added partner's subvention $\phi\pi$ against the spoils dilution $\Pi/n - \Pi/(n + 1)$ and the programming strain. When σ_η is large $p(n, \rho)$ climbs steeply from a low base, so $\bar{g}$ outruns $\underline{g}$ over a range above the minimal-winning size and the gains exceed the dilution: the coalition is oversized. When σ_η is small p is near its ceiling already, the monitoring gain is slight, and the dilution and the rising $\underline{g}$ bind, contracting the coalition to the minimal programmable winning set. The size is an interior balance, struck above minimal-winning when the signal is noisy.

11 Formation and coalition-proofness

Endogenise the coalition. A formateur assembles a set that is winning, programmable, and monitorable; with three or more parties a further test enters, since a sub-coalition may break away and govern alone. The proposed coalition is stable only if no winning subset can do

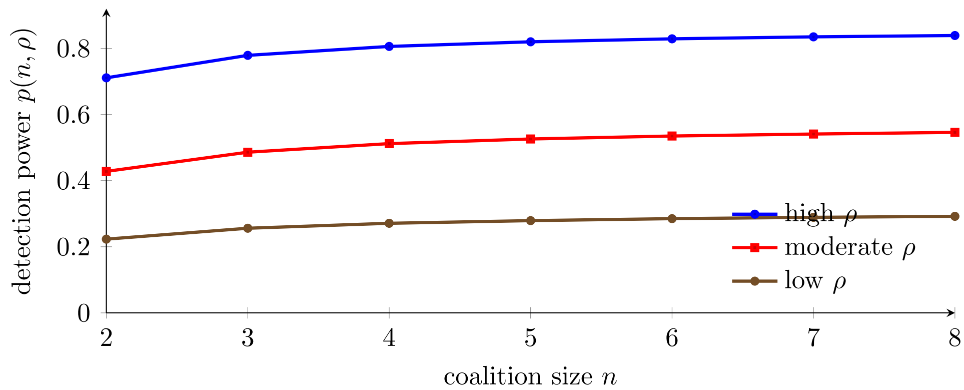

Figure 1: Monitoring sharpens with size. The detection power $p(n, \rho) = 2\Phi\big(\frac{\kappa}{2\sigma_\eta}\sqrt{(n-1)/n}\big) - 1$, the chance a lone deviator is named by the contrast (6), rises with the number of partners n and with the correlation ρ of their fortunes. A coalition monitors better the wider it is, and better still when its members rise and fall together.

better apart, the coalition-proofness of Bernheim, Peleg, and Whinston (1987) laid over the renegotiation problem of Farrell and Maskin (1989).

Proposition 6 (Coalition-proof governments)**.** *A winning, programmable, monitorable coalition C is stable against breakaway if and only if every winning subset $C' \subset C$ either (i) is itself unable to discipline – too few partners to attribute or too spread to programme once the others are gone – or (ii) earns its members less in spoils and policy than C does. Two forces hold a coalition together against its own subsets: the spoils of size, since a smaller government commands a smaller subvention, and monitorability, since a subset that sheds partners may lose the cross-section it needs to police itself. An oversized coalition is sustained when its surplus partner is essential for attribution: dropping it would win just as surely but see far worse.*

Coalition-proofness thus pulls toward the minimal programmable subset, while monitoring pulls toward breadth; the stable government balances the two. Where minimal-winning theory treats every surplus partner as a puzzle to be explained away, the repeated account treats it as the price of discipline, and predicts that the surplus partners which survive the breakaway test are those whose polls the coalition cannot do without.

12 Empirical implications

The model speaks to data in terms the one-shot theory does not reach, and, because its central objects are a line, a partner's position, and its advocacy of the line or drift from it, much of what it asks for can be read from text.

Five hypotheses. The account yields predictions that can be taken to data. H1 (Composition). Holding the seat arithmetic and ideology fixed, parties whose polls are more correlated coalesce more readily: monitorability, not minimal-winning size alone, selects partners. H2 (Size). The incidence of surplus partners rises with the idiosyncratic, party-specific variance of polls, as in volatile or fragmented systems, new democracies, and thin polling, because there the marginal partner is taken on for the monitoring its poll supplies, not for a majority already in hand. H3 (Survival, relative not absolute). The hazard of termination loads on the cross-partner divergence of fortunes, not the coalition's common level: a partner slipping against its fellows imperils a government more than a downturn that sinks them together. H4 (Form). Conditional on a crisis, the chance it ends in one partner's exit rather than the whole government's fall rises with the partners' poll correlation: legible coalitions reshuffle, illegible ones collapse. H5 (Blameless terminations). Some governments fall with no antecedent divergence at all, on noise no partner authored, which is the model's signature and the prediction the opportunity account of termination cannot make.

Measurement, and a role for text. The model's latent objects map onto observables, several of them textual. The agreed line m is itself a negotiated document, the coalition agreement, which is a direct reading of the programme at formation. Each partner's position x_i, and the span of the coalition that fixes its programmability and survival threshold, are recoverable from manifestos, whether hand-coded in the manner of the Comparative Manifesto Project or scaled from the raw text (Laver, Benoit, and Garry, 2003; Slapin and Proksch, 2008). The action the model cannot see, a partner's advocacy of the line versus its drift toward its own ideal, becomes observable as the distance between a partner's ongoing speech and the agreed line: a partner that holds the line speaks close to the coalition agreement, one that breaks ranks speaks away from it, and scaling that distance through a run of parliamentary text (Grimmer and Stewart, 2013) yields a measured drift series for each partner over the life of a government. The correlation ρ is read directly, from the covariance of partners' polling series.

The data. Each construct has a cross-national source, and the model asks for nothing not already, in principle, assembled. Cabinet durations and the manner of each termination, the dependent variable, are catalogued in the comparative cabinet datasets, from Strøm, Müller, and Bergman (2008) to the ParlGov and European Representative Democracy collections, which record a government's formation, composition, and end. Parties' positions x_i and the coalition's span come from the Manifesto Project (Volkens et al., 2021) or from text-scaling (Laver, Benoit, and Garry, 2003; Slapin and Proksch, 2008); the agreed line m is the coalition agreement, a growing share of them digitised; partners' drift is read from parliamentary speech, now available at scale in cross-national corpora such as ParlSpeech (Rauh and Schwalbach, 2020); and the correlation ρ, with the relative-decline measure of H3, is built from national polling series.

An estimation. Textual drift is the noisy signal of defection the model says a coalition punishes, and the pieces compose into a competing-risks hazard of termination in the manner of

Diermeier and Stevenson (1999), with the model's covariates: each partner's drift from the agreed line, the cross-partner divergence of drift and of polls, the correlation ρ, and the interactions drift $\times$ ρ and divergence $\times$ ρ the theory makes central. H3 predicts a positive loading on divergence and a null on the common level; H4, that the discretionary risk – a partner replaced, the cabinet reshuffled – rises with ρ while the electoral risk of dissolution does not; H5, a mass of terminations with drift near zero.

Distinguishing the model. The tests that separate this account from its rivals are built into these covariates. Against the opportunity theory of Diermeier and Stevenson (1999, 2000), in which a critical event ends a government because dissolution has become wise, the model predicts terminations not warranted in hindsight, namely blameless collapses on noise (H5), and crises keyed to a partner's relative, not the coalition's absolute, standing (H3). Against minimal-winning accounts of composition (Riker, 1962), it predicts surplus partners that are essential for monitoring, dropping out of the coalition where the signal is clean (H2). And against strategic-timing accounts (Lupia and Strøm, 1995), in which an incumbent dissolves to seize an electoral advantage, it predicts dissolution on a partner's adverse relative drift even when the common tide is favourable, a government brought down while it could have won.

Feasibility. The estimator's power at the sample sizes the cross-national cabinet record affords, and the design's recovery of the predicted contrasts (targeted exits loading on relative divergence and flat in the common level, collective exits loading on the level, the form sorting on ρ, and a mass of blameless terminations), are matters for the empirical companion, of which the assembled panel is the subject; we draw out here only what the theory asks that companion to find.

13 An illustration: the traffic-light coalition

The model's signature, discipline by the targeted exit of the partner whose drift its fellows can read against their own fortunes, appears in the German *Ampel* of 2021–24. Three parties governed: the Social Democrats, the Greens, and the Free Democrats. Through 2023–24 the Free Democrats' support slipped toward and below the five-per-cent threshold while their partners held. This was a decline relative to the coalition, not a common tide, and the slippage the model says imperils a government. Their response was the deviation discipline must catch: pressed to re-differentiate as their standing fell, they advanced a more open fiscal line of their own, culminating in a circulated economic-policy paper at odds with the government's programme. That drift was legible, a position set down in text, a measurable distance from the coalition agreement, the observable of Section 12, and the crisis resolved not in a symmetric dissolution but in the expulsion of the named partner: the Chancellor dismissed the Free Democrats' finance minister, the party left office, and a Social Democratic–Green minority governed on to an early election. A coalition that could see which partner had drifted disciplined that partner rather than the government as a whole, the targeted capture of Section 6 in the form the model predicts, with the remaining partners content to govern without the one the signal had named. The case is a single illustration, not a test, but it shows the model's distinctive objects, relative decline, a drift left in the record, and a discipline aimed rather than collective, standing together in one government's fall.

14 Concluding remarks

We have taken a coalition to be not a one-shot bargain but a repeated disciplinary structure. What a coalition can see governs what it can be. Because a government holds the whole vector of its partners' polls, it can net out the national tide and use the residual that identifies a defector. Because a wider coalition estimates that tide more sharply, the power to monitor grows with size. Coalitions are assembled, and sometimes oversized, for legibility rather than minimality. The same capacity governs survival: a coalition that uses the residual survives a national swing that would collapse a coalition relying only on the aggregate. Governments fall, on the path, on coalitions no member chose to break. The classical theory asks who is in and who gets what. This one asks who can be disciplined, and answers with the monitoring structure of electoral fortunes. The single party of Dewan (2026) is the case $n = 1$, where there is no cross-section to read and no tide to net out. Coalitions are where attribution becomes possible, and where it becomes central to the politics.

This is a small step, and a first-shot. We have kept to the theory and the testable shape it takes, and left the estimation to the companion the data deserve: the agreed line read against the covariance of the polls, advocacy and drift measured in the parliamentary record. What the model offers meanwhile is a way to see a government's fall not as the running-out of a clock but as a punishment a coalition saw coming and could not, in the end, forgive.

Appendix: Proofs

Write Φ, ϕ for the standard normal cdf and density. The per-period payoffs in the cooperative and punishment phases are $w_C > w_P$, with the deterrent $\Delta = V_C - V_P$ and phase values

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

so that $\Delta = (1-\delta^T)\big(V_C - w_P/(1-\delta)\big)$ rises in T to the grim bound $\Delta_{\max} = V_C - w_P/(1-\delta)$, exactly as in Dewan (2026); the optimal-trigger and comparative-statics arguments there apply unchanged to (4) and are not repeated.

Proof of Lemma 1. Under universal advocacy $y_i - \pi_i = u + \eta_i$, so $\frac{1}{n}\sum_j (y_j - \pi_j) = u + \bar\eta$ and

$r_i = (u+\eta_i)-(u+\bar\eta) = \eta_i - \bar\eta$; the tide cancels for every realisation of u. As a linear combination of independent normals, r_i is normal with mean 0 and $\mathrm{Var}(r_i) = \mathrm{Var}(\eta_i) - 2\,\mathrm{Cov}(\eta_i, \bar\eta) + \mathrm{Var}(\bar\eta) = \sigma_\eta^2 - 2\sigma_\eta^2/n + \sigma_\eta^2/n = \sigma_\eta^2(1-\frac{1}{n})$. If only k deviates, $y_k - \pi_k = u - \kappa + \eta_k$ and the mean over j is $u - \kappa/n + \bar\eta$, giving $\mathbb{E}[r_k] = -\kappa + \kappa/n = -\kappa(1-\frac{1}{n})$ and, for $i \neq k$, $\mathbb{E}[r_i] = \kappa/n$. The difference $r_k - r_i$ has mean $-\kappa$, and $r_k - r_i = (\eta_k - \eta_i)$ in distribution up to the common $\bar\eta$ shift, with $\mathrm{Var}(\eta_k - \eta_i) = 2\sigma_\eta^2$, free of n and of σ_u. □

Proof of Lemma 2. Under universal advocacy the leave-one-out average is $\frac{1}{n-1}\sum_{j\neq i}(y_j - \pi_j) = u + \frac{1}{n-1}\sum_{j\neq i}\eta_j$, so $c_i = (u+\eta_i) - \big(u + \frac{1}{n-1}\sum_{j\neq i}\eta_j\big) = \eta_i - \frac{1}{n-1}\sum_{j\neq i}\eta_j$: the tide cancels for every u. By independence the mean is 0 and $\mathrm{Var}(c_i) = \sigma_\eta^2 + \frac{1}{(n-1)^2}(n-1)\sigma_\eta^2 = \sigma_\eta^2\frac{n}{n-1}$. If i alone deviates, $y_i - \pi_i = u - \kappa + \eta_i$ while the others are unchanged, so c_i shifts by $-\kappa$, mean $-\kappa$ and the same variance. At $n = 2$ the average is the single other party and $c_i = (y_i - \pi_i) - (y_j - \pi_j)$, variance $2\sigma_\eta^2$, the difference of Dewan (2026). □

Proof of Proposition 1. By Lemma 2, $c_i \sim N(m, \sigma_\eta^2 n/(n-1))$ with $m = 0$ under universal advocacy and $m = -\kappa$ when i alone deviates. The test names i when $c_i < -c$, so $\Pr[\text{name } i] = \Phi\big((-c-m)\sqrt{(n-1)/n}/\sigma_\eta\big)$, which is π_0 at $m = 0$ and π_1 at $m = -\kappa$, the expressions (7). As $\kappa - c > -c$, $\pi_1 > \pi_0$, and $\pi_1 - \pi_0$ is increasing in $(\kappa/\sigma_\eta)\sqrt{(n-1)/n}$, hence in κ/σ_η (so in ρ for fixed σ_u) and in n (as $\sqrt{(n-1)/n}$ rises). At $\sigma_\eta \to 0$ ($\rho \to 1$), $\pi_1 \to 1$ and $\pi_0 \to 0$; at $\sigma_\eta \to \infty$ ($\rho \to 0$), both tend to $\frac{1}{2}$ and $\pi_1 - \pi_0 \to 0$.

Renegotiation-proofness. Expelling the named party hands the line and the spoils to the remaining coalition; where that remainder is winning, its members strictly prefer governing without the named party – a larger share each and a line nearer their ideals – so they will not readmit it during the phase, the continuation is not Pareto-dominated for those who enforce it, and the scheme is weakly renegotiation-proof in the sense of Farrell and Maskin (1989), the coalitional counterpart of the capture in Dewan (2026). (Where the named party is pivotal to the majority, the punishment is instead the line swinging to the remainder's preferred point, which they likewise prefer.)

Incentives. Let V_C be the value of continued cooperation and $V^{\text{out}} < V_C$ that of the expelled party during the phase. A member i weighing a one-shot deviation in a cooperative period collects $g_i - \phi\kappa$ now and moves c_i's mean from 0 to $-\kappa$, raising the chance it is the expelled party from π_0 to π_1; its continuation falls by at least $\delta(\pi_1 - \pi_0)(V_C - V^{\text{out}})$. Advocacy is therefore incentive-compatible whenever $\max_{i\in C} g_i - \phi\kappa \leq \delta(\pi_1 - \pi_0)(V_C - V^{\text{out}})$, binding on the most-tempted partner. As $\rho \to 1$ the deviator is expelled with certainty and the innocent never, so the blameless collapse vanishes; as $\rho \to 0$ the contrast is uninformative and discipline reverts to the aggregate trigger of Section 4. □

Proof of Proposition 2. The coalition's test for "member k deviated" against "no deviation" is a likelihood-ratio test on the residual vector r. By Lemma 1 the deviation shifts r_k by $-\kappa(1-\frac{1}{n})$ against per-comparison noise $\sqrt{2}\,\sigma_\eta$; the test's power $p(n, \rho)$ is the probability the true deviator's residual is the most negative by the critical margin, a function strictly increasing in the separation $\kappa(1-\frac{1}{n})/\sigma_\eta$ – hence in n and, for fixed σ_u, in ρ (since $\rho\uparrow$ means $\sigma_\eta\downarrow$). As $\sigma_\eta\to 0$ ($\rho\to 1$) the separation diverges and $p\to 1$: the deviator is named with certainty and an asymmetric punishment falls on it alone. As $\sigma_\eta\to\infty$ ($\rho\to 0$) the separation vanishes, r is uninformative, and the only deterrent left is the aggregate trigger with its collective collapse. A targeted punishment of expected per-deviation cost proportional to p relaxes (4); where p is high the constraint holds for larger $\max_i g_i$, where p is low it holds only for small $\max_i g_i$, the collective collapse bearing the deadweight $\delta q_0\Delta$. □

Proof of Proposition 3. With $g_i = g(|m-x_i|)$ and g increasing, the binding temptation is $\max_{i\in C} g(|m-x_i|) = g\big(\max_{i\in C}|m-x_i|\big)$, and $\max_{i\in C}|m-x_i| = \max\{m-\underline{x},\ \bar{x}-m\}$ with $\underline{x} = \min_i x_i$, $\bar{x}=\max_i x_i$. This is minimised where the two arguments are equal, at $m^* = \frac{1}{2}(\underline{x}+\bar{x})$, giving residual temptation $g\big(\frac{1}{2}(\bar{x}-\underline{x})\big)$ – the centre of the coalition's span (the midpoint of ideals in the symmetric case). Unity on the best line is sustainable iff

$$g\big(\tfrac{1}{2}(\bar{x}-\underline{x})\big) - \phi\kappa \ \le\ \delta\, D^*\,\Delta_{\max},$$

where D^* is the detection power of the optimal scheme (the aggregate q_1-q_0, or its residual-augmented value when attribution is used). The right side is independent of the spread $\bar{x}-\underline{x}$ while the left strictly increases in it (with $g(0)-\phi\kappa<0$ at zero spread), so a unique $\bar{D}>0$ solves the equality and unity fails for $\bar{x}-\underline{x}>\bar{D}$. Disciplinability requires in addition $D^*>0$, i.e. a detection power – from ρ, n through attribution or from the aggregate trigger – sufficient that the inequality can hold; the viable set is the intersection. □

Proof of Proposition 4. Under the residual scheme the statistic governing punishment is r, which by Lemma 1 is distributed independently of u; the probability of a (false or true) punishment in a period therefore does not depend on σ_u, and a realisation of the tide u, however large, leaves the residual law unchanged – the coalition reads every partner's standing as intact and does not punish. When ρ is low the residual is uninformative (its separation $\kappa(1-\frac{1}{n})/\sigma_\eta\to 0$ as $\sigma_\eta\to\infty$), so the best scheme falls back on the aggregate statistic Y, with $\mathrm{Var}(Y) = n^2\sigma_u^2 + n\sigma_\eta^2$ by (3); the false-collapse probability $q_0 = \Phi\big((\hat{Y}-\mathbb{E}Y)/\sqrt{\mathrm{Var}(Y)}\big)$ at a fixed deterrence target rises with σ_u and with n, and a low realisation of u drives $Y<\hat{Y}$ and triggers collapse though $d=0$. Hence survival is independent of the tide where attribution is available (ρ high) and falls to it where attribution fails (ρ low), with no coalition ever choosing the aggregate when the residual would serve. □

Proof of Proposition 5. Let the coalition use the residual scheme with fallback collapse, and write its sustainable temptation ceiling as $\bar{g}(n) = \phi\kappa + \delta\, D^*(n)\,\Delta_{\max}$, where $D^*(n)$ is increasing in n through $p(n,\rho)$ (Proposition 2). Adding the partner with ideal nearest the existing span raises the required temptation $\underline{g}(n) = g\big(\frac{1}{2}(\bar{x}_n-\underline{x}_n)\big)$ weakly, by widening the span, while raising

$\bar{g}(n)$ through improved detection; discipline is sustainable at size n iff $\underline{g}(n) \leq \bar{g}(n)$. When σ_η is large, $p(n,\rho)$ and hence $\bar{g}(n)$ rise steeply in n from a low base, so a range of n above the minimal-winning size first makes $\underline{g} \leq \bar{g}$ feasible: the discipline-optimal coalition is oversized. When σ_η is small, p is near one even at small n, $\bar{g}(n)$ is already slack, and only the rising $\underline{g}(n)$ binds, so the optimum is the smallest programmable winning coalition. Continuity in σ_η gives an intermediate regime with an interior optimal size. □

Proof of Proposition 6. A winning subset $C' \subset C$ can profitably break away iff, governing alone, it (a) can sustain discipline – $\underline{g}(C') \leq \bar{g}(C')$ in the sense of Proposition 5, evaluated at C''s own spread, correlation, and size – and (b) delivers each of its members more than C does. The office return is ϕ times coalition standing, which is smaller for the smaller C' by the loss of the departed partners' baselines and subvention, so (b) fails unless C''s tighter programme (a line closer to its members' ideals, lower policy loss) outweighs the forgone spoils. And (a) fails when C' drops a partner essential for attribution: by Lemma 1, $D^*(C') < D^*(C)$, and if the lost detection takes $\bar{g}(C')$ below $\underline{g}(C')$ the subset cannot police itself. C is stable iff for every winning $C' \subset C$ at least one of (a), (b) fails, which is the stated condition; an oversized C is stable when its surplus partner is essential for (a), so that every majority subset dropping it is undisciplinable. Read as an induction on coalition size – each breakaway subset judged stable by the same test – the condition delivers coalition-proofness in the sense of Bernheim, Peleg, and Whinston (1987) rather than mere immunity to a single round of breakaway. □

References

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

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

Baron, D. P., and J. A. Ferejohn (1989). Bargaining in Legislatures. *American Political Science Review* 83(4), 1181–1206.

Bernheim, B. D., B. Peleg, and M. D. Whinston (1987). Coalition-Proof Nash Equilibria I: Concepts. *Journal of Economic Theory* 42(1), 1–12.

Dewan, T. (2026). Splits as Price Wars: Party Discipline under Imperfect Monitoring. Working paper.

Diermeier, D., H. Eraslan, and A. Merlo (2003). A Structural Model of Government Formation. *Econometrica* 71(1), 27–70.

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

Diermeier, D., and A. Merlo (2000). Government Turnover in Parliamentary Democracies. *Journal of Economic Theory* 94(1), 46–79.

Diermeier, D., and R. T. Stevenson (1999). Cabinet Survival and Competing Risks. *American Journal of Political Science* 43(4), 1051–1068.

Diermeier, D., and R. T. Stevenson (2000). Cabinet Terminations and Critical Events. *American Political Science Review* 94(3), 627–640.

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

Friedman, J. W. (1971). A Non-cooperative Equilibrium for Supergames. *Review of Economic Studies* 38(1), 1–12.

Fudenberg, D., and E. Maskin (1986). The Folk Theorem in Repeated Games with Discounting or with Incomplete Information. *Econometrica* 54(3), 533–554.

Fudenberg, D., D. Levine, and E. Maskin (1994). The Folk Theorem with Imperfect Public Information. *Econometrica* 62(5), 997–1039.

Gamson, W. A. (1961). A Theory of Coalition Formation. *American Sociological Review* 26(3), 373–382.

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.

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.

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.

Laver, M., and K. A. Shepsle (1996). *Making and Breaking Governments: Cabinets and Legislatures in Parliamentary Democracies.* Cambridge University Press, Cambridge.

Lupia, A., and K. Strøm (1995). Coalition Termination and the Strategic Timing of Parliamentary Elections. *American Political Science Review* 89(3), 648–665.

Merlo, A. (1997). Bargaining over Governments in a Stochastic Environment. *Journal of Political Economy* 105(1), 101–131.

Rauh, C., and J. Schwalbach (2020). The ParlSpeech V2 Data Set: Full-Text Corpora of 6.3 Million Parliamentary Speeches in the Key Legislative Chambers of Nine Representative Democracies. Harvard Dataverse.

Riker, W. H. (1962). *The Theory of Political Coalitions*. Yale University Press, New Haven.

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.

Strøm, K. (1990). *Minority Government and Majority Rule*. Cambridge University Press, Cambridge.

Strøm, K., W. C. Müller, and T. Bergman (eds.) (2008). *Cabinets and Coalition Bargaining: The Democratic Life Cycle in Western Europe*. Oxford University Press, Oxford.

Volden, C., and C. J. Carrubba (2004). The Formation of Oversized Coalitions in Parliamentary Democracies. *American Journal of Political Science* 48(3), 521–537.

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

Warwick, P. (1994). *Government Survival in Parliamentary Democracies*. Cambridge University Press, Cambridge.