

Contestable Leadership: Party Leaders as Principals *and* Agents*

Matias Iaryczower

Assistant Professor of Economics and Political Science, Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125, USA;
miaryc@hss.caltech.edu

ABSTRACT

This paper examines the institutional determinants of discipline in legislative parties. The model formalizes the tradeoff between resources at the leader's discretion, and the leader's need to maintain a minimum level of support to continue leading. The value of the leader's promises of future benefits is here endogenously determined by the backbenchers' beliefs about the extent of support to the leader among other party legislators. Rewards that can be distributed publicly and on the spot are effective tools to coordinate beliefs about the stability of the leader, and thus also increase the value of the leader's promises of future benefits. These spot resources are in fact necessary for the leader to be powerful: without them, the leader can use promises of future benefits to sway members' behavior only if a majority of the party agrees (*ex ante*) with the leader's preferred position in the first place.

Keywords: Party discipline; legislatures; leadership; vote buying; global games.

JEL Codes: D72, D78, C72

The predominant view among scholars studying European and Latin American legislative party organizations grants significant power to party leaders. Based on leaders'

* This paper is a revised version of the first chapter of my PhD dissertation. I am very grateful to David Levine, Andy Atkeson, Hal Cole, Christian Hellwig, Leeat Yariv, and Juliana Bambaci for their helpful discussions and suggestions. I am also thankful to numerous seminar participants at Caltech, Chicago, Illinois at Urbana-Champaign, Maryland, U. Penn, Princeton, Rice, Stanford, UC Berkeley, UCSD, and Washington University in St. Louis, as well as to many others at UCLA, for their useful comments to previous versions of this paper.

control of nominations, committee assignments, and other resources at their disposal, the leadership is able to discipline backbenchers to follow the party line.¹ On the other side of the spectrum, the currently predominant theories of legislative leadership in the United States conceive leaders as agents of the rank and file (Sinclair 1983, Kiewiet and McCubbins 1991, Cox and McCubbins 1993, Aldrich and Rohde 1997).

The above characterization suggests two worlds apart, where leaders are *principals* in some environments, and *agents* in others. Power relations, however, are seldom black and white. On the one hand, party leaders rarely have unconditional control of party resources. Instead, their authority relies on an implicit contract with backbenchers that grants them legitimacy: leaders manage party resources conditional on the support of the organization. Nowhere is this so apparent as in the US Congress, as it was recognized early by Jones (1968), Sinclair (1983), and Calvert (1987) among others. But the fact that effective leadership requires the support of the followers is nonetheless true in other latitudes as well (McKenzie 1964, Panebianco 1988, Bowler *et al.* 1999a, Myerson 2005). The difference is one of degree, not of nature.

On the other hand, the *party* is a collective, heterogeneous, principal. Under these circumstances, the leader does not merely solve problems of coordination (rallying support around one of possibly many acceptable alternatives), but instead *de facto* resolves conflict, leading members to pursue policies with less than total support within the party.² A rogue agent, equipped with resources to discipline his/her own principals, makes for an uncomfortable principal agent theory. A possible way out of this conundrum is the theory of conditional party government of Aldrich and Rohde (Rohde 1991, Aldrich 1995, Aldrich and Rohde 1997). In their view, the rank and file does not delegate powers to the leadership unless their views are sufficiently homogeneous: "If there is much diversity in preferences within a party, a substantial portion of the members will be reluctant to grant strong powers to the leadership, or to resist the vigorous exercise of existing powers, because of the realistic fear that they may be used to produce outcomes unsatisfactory to the members in question" (Aldrich and Rohde 1997). In the extreme, however, the theory of conditional party government does not provide much of a way out of the conundrum after all (Krehbiel 1998). The problem is that it is exactly in those circumstances in which the leader is granted strong powers according to the conditional party government hypothesis, that there is not much need for power to be exercised in the first place. Party activity might then be visible, but not too important when compared with what would have happened in the absence of partisan organizations. Away from the extremes of a perfectly homogeneous or very conflictual party, we are left with specifying how the collective action problem of the collective (heterogeneous) principal is resolved.

¹ See any of the contributions in Morgenstern and Nacif (2002) or Bowler *et al.* (1999b).

² As Kiewiet and McCubbins (1991) argue, "The key determinant of the desirability of checks within the structure of party leadership is the degree of homogeneity in the policy preferences of the membership . . . when the party caucus is riven by serious policy disputes, there is more support for checks. Without them, one faction, upon gaining control of the machinery of leadership, might pursue policies that are anathema to another faction, thereby weakening or even splintering the party" (p. 54).

In this paper, I build on the notion of conditional party government, introducing formally the concept of contestable leadership.³ I also emphasize here, however, the collective action problem of the collective principal. I start from the premise that resisting the exercise of existing powers (or removing existing powers) is not the choice of a single backbencher, but instead a collective problem among a heterogeneous group of backbenchers. I argue that the key element delimiting the power of the leader is how this impure coordination problem among backbenchers is resolved. From this perspective, the alternative views expressed in the leaders-as-principals and leaders-as-agents traditions can be taken to represent opposite assumptions of how difficult it is for the rank and file to coordinate to effectively check the leader's power: while this coordination is precluded outright in the world of the *iron law of oligarchy* (Michels 1962), it is assumed to work without friction in the conditional party government framework of Aldrich and Rohde, and in the cartel theory of Cox and McCubbins.

The measure of power is party discipline. The concept of party discipline refers to the power of the party leadership to induce cohesive or conformist behavior of backbenchers even in spite of significant internal dissent (Krehbiel 1993, Cox and McCubbins 1993, Tsebelis 1995). I implement the concept of discipline by considering the difference in equilibrium cutpoints (the ideal policy of the most left-winged backbencher voting for the right-wing alternative in equilibrium) in whip and no-whip votes. I argue that this measure satisfies several desirable properties and is easily implementable in analysis of roll call voting (Snyder and Groseclose 2000, MaCarty *et al.* 2001). I distinguish between payments that can be distributed on the spot to both members of the party and the opposition (*pork*), and promises of future partisan benefits, such as nomination to party lists. Thus in this context, discipline is a measure of the power of partisan promises to induce behavior different than in a non-partisan benchmark.

The key effect of a leadership replacement in the model is its effect on the allocation of promises of future benefits, which as opposed to spot payments can only be delivered if the leader retains the command of the party. The central theme that unifies the results of the paper follows directly from this fact. Since promises are conditional on the stability of the leader, their value is not exogenously given, but instead endogenously determined by backbenchers' beliefs about the extent of support to the leader among other party legislators. How these beliefs are formed determines how the collective action problem among backbenchers opposing the leader is resolved, and thus ultimately how powerful the leader is *vis-à-vis* the collective principal.

I show that rewards that can be distributed publicly and on the spot are effective tools to coordinate beliefs about the stability of the leader, and thus also to increase the value of the leader's promises of future benefits (and with it party discipline). These spot resources are in fact necessary for the leader to be powerful: without them, the leader can use promises of future benefits to sway members' behavior only if a majority

³ For other formal models that incorporate challenges to political leaders in other contexts, see Myerson (2005) and Bueno de Mesquita *et al.* (2003). For other recent articles building on the notion of conditional party government, see Lebo *et al.* (2007), Patty (2008), and Volden and Bergman (2006).

of the party agrees (*ex ante*) with the leader's preferred position in the first place. This result complements the conventional wisdom prevailing in the discipline regarding the impact of nomination power on party discipline. The general moral is that promises of future benefits will be relevant to alter voting behavior only if party members believe that the leader has a strong hold to the reins of power. With no resources to distribute on the spot, the leadership of a majoritarian institution is only conditionally stable, and the leader merely an agent of the majority of the party. When instead the leader has access to resources that she can guarantee to recipients independently of the resolution of strategic uncertainty (i.e., pork), it becomes common knowledge among backbenchers that the leader is likely to retain the command of the party. These beliefs in turn increase the power of partisan resources. In a nutshell, discipline requires that challenging the party line is perceived to be difficult among backbenchers, and something other than promises is needed to pin down these beliefs.

The model delivers two additional empirical implications. First, I show that under mild assumptions, party discipline is increasing in the heterogeneity of preferences among backbenchers. I also consider the allocation of pork between the party and the opposition. Intuitively, a leader who is strong within the party does not need to deploy resources to obtain support within the party, and can instead use pork to increase support within the opposition. I show that this is true even if the leader does not care about the leadership position *per se*, but only about policy i.e., that more vulnerable leaders will allocate a higher proportion of pork to buy members of their own party *vis-à-vis* the opposition.

THE MODEL

There are three types of agents in the model: a *party leader*, a continuum of *backbenchers*, with total size 1, and a continuum of size β of *opposition legislators*. Backbenchers and opposition legislators integrate a legislature, which chooses between two given policy alternatives q and x in \mathbb{R} , $q < x$, by simple majority voting.⁴ The party leader is endowed with resources that allow him/her to influence legislators' voting behavior, and cares exclusively about policy outcomes. In particular, we assume without loss of generality that the leader's ideal policy is to the right of the midpoint between x and q , and therefore prefers the policy outcome to be x .

The essence of the argument is that the disciplining power of promises of future benefits depends fundamentally on whether backbenchers perceive that the leader's allocation of resources is likely to stay firm or be reversed as a result of dissent within the party. The model therefore requires three basic elements: (i) *strategic uncertainty* for backbenchers, (ii) a distinction among resources available to the leader between those that the leader can *guarantee* to the prospective recipients, and those that depend on the resolution of the strategic uncertainty, and (iii) a model of the *contestable* nature of the

⁴ With fixed alternatives, we can as well take our policy space to be \mathbb{R}^n . Similarly, as it will be apparent soon, nothing here depends on the voting rule being simple majority.

leadership position, linking the strategic uncertainty at the level of backbenchers in (i) with the value of offers in (ii).

To allow strategic uncertainty at the level of backbenchers in equilibrium, we introduce incomplete information about their payoffs. Backbenchers' payoffs are determined by (i) *rents* they can extract from the party leadership, and (ii) a *position-taking* cost given by the distance between their constituents ideal policy $\theta_i \in \mathbb{R}$, and the policy they voted for in the legislature, $x_i \in \{q, x\}$. Backbenchers have quasilinear preferences. The position-taking preferences of backbencher i are represented by a decreasing, concave, and differentiable utility function $u(|x_i - \theta_i|)$, and rents t_i enter linearly into their utility function, so that $U(x_i, t_i; \theta_i) = u(|x_i - \theta_i|) + t_i$. It will be convenient to define — taking the pair (q, x) of policy alternatives as given — the function $v(\theta_i) \equiv u(|q - \theta_i|) - u(|x - \theta_i|)$. The value $v(\theta_i)$ denotes the net gain of voting for q instead of x for a backbencher i with ideal policy θ_i . Note that $v(\cdot)$ is strictly decreasing, that $v(\theta_i) = 0$ at $\theta_i = \frac{x+q}{2}$, and that $|v(\cdot)|$ is symmetric around this point. I will also assume that $|v(\cdot)|$ is convex and therefore unbounded, assuring that the position-taking cost dominates for a sufficiently extreme individual.⁵ The ideal policy of each backbencher, θ_i , is private information, but correlated with that of the other backbenchers.⁶ Specifically, assume that the ideal policy of backbencher i is given by $\theta_i = \theta + \varepsilon_i$, where the common component θ is drawn from a $N(\theta_0, \eta^2)$ distribution, the idiosyncratic component ε_i is iid, and drawn from a $N(0, \sigma^2)$ distribution, and both θ and ε_i are unobservable. This uncertainty about the preferences of fellow party members allows us to retain strategic uncertainty at the level of backbenchers even in a Nash equilibrium of the voting game.

The second pillar of the model is the distinction, among resources available to the leader, between those that the leader can guarantee to the prospective recipients, and those that depend on the resolution of strategic uncertainty. To focus on backbenchers' collective action problem of whether to follow or rebel against the leader, I restrict to contracts of the simplest possible type, where *offers* (but not necessarily realized payments) are contingent on the action of each individual legislator only, and symmetric among legislators of the same party.⁷ I distinguish here between *pork*, and *promises* of future partisan benefits. Offers of pork, of total value R , are firm conditional offers that

⁵ For all results but Remark 2 and Proposition 4, where I explicitly use the convexity of $|v(\cdot)|$ in the proof, it is enough to require $|v(\cdot)|$ to be unbounded. A sufficient condition for $|v(\cdot)|$ to be convex is that $u''' < 0$. This is satisfied, for example, by quadratic utility functions, and by CRRRA utility functions with arguments being decreasing functions of distance.

⁶ With minor modifications, we could assume that the preferences of a fraction of *partisan* backbenchers are public information, and carry out the analysis with respect to the remaining fraction of *moderate* backbenchers.

⁷ These rule out differentiated payments or promises to legislators of the same party, as well as more complex mechanisms that could possibly depend on aggregate voting patterns. In particular, it also rules out offers that use exhaust resources *ex post*. Assume that the leader keeps all residuals from offers that were not accepted. For models that allow offers to depend on aggregate voting results, see Dal Bo (2007), Dekel *et al.* (2006), and Morgan and Vardy (2006). Morgan and Vardy (2006), in fact, show that when agents care solely about their vote, it is better for a vote buying principal to use vote buying contracts that depend on the realization of the vote, rather than on the realization of the aggregate voting outcome. For alternative vote buying models that are richer in other respects,

can be extended to both backbenchers and opposition legislators, and executed on the spot following a vote, i.e., independently of the actions of other players, a backbencher (opposition legislator) receives r (r_o) for sure when voting in favor of x , and zero otherwise. The party leader chooses $r \in [0, R]$ subject to the (*ex ante*) budget constraint $r_o\beta + r = R$. Promises of future partisan benefits, instead, are unsettled conditional offers, whose delivery depend on the degree of support to the leader among backbenchers. The strategic choice of the leader with regard to these promises of future partisan benefits is not one of allocation, but of whether to put these resources in play or not, i.e., the leader has gatekeeping power over these resources. If the leader does not, promises do not enter into backbenchers' payoffs, and the net payoff of voting for x for backbencher i is then given by $\Pi_n(\theta_i) = r - v(\theta_i)$. If the leader does, he/she induces a *whip* vote. The cost of employing promises of future party resources to support the party line is that doing so triggers a potential challenge to the party leader (the third basic element of the model). If the support of backbenchers to the party line is sufficiently high, then promises are delivered as promised by the leader. If, however, the mass of backbenchers in the leader's coalition, denoted by Γ , does not reach a minimum threshold μ ($\mu \leq 1/2$), then the conditional allocation of future resources promised by the leader is reversed, and backbenchers voting for q receive partisan benefits e , while those voting for x receive zero. As a result, the net expected payoff of voting for x for backbencher i in a whip vote is then $\Pi_m(\theta_i) = r + e[1 - 2\Pr(\Gamma < \mu|\theta_i)] - v(\theta_i)$.

To define strategies and equilibrium, it remains to specify the problem of opposition legislators, which I keep here as simple as possible. To do so, assume that opposition legislators have policy preferences $u(\cdot, \theta_i^o)$, with ideal policies distributed according to a known cdf $G(\cdot)$. This implies, in particular, that the proportion of opposition legislators with ideal policy below some number z is public information.⁸ Since from the budget constraint for offers of pork $r_o = \frac{R-r}{\beta}$, the mass of legislators in the opposition voting for x for any offer r to backbenchers is given by $[1 - G(v^{-1}((R-r)/\beta))]\beta$. I now define strategies and equilibrium considering the best responses of opposition legislators as part of the environment, and excluding them outright from the set of players. The timeline of the game consists of three stages. In Stage 1, nature chooses a realization of the unobservable random variables θ and $\{\varepsilon_i\}$, and each backbencher i privately observes his/her ideal policy $\theta_i = \theta + \varepsilon_i$. The party leader receives no such private signal (all his/her information is public, and incorporated in the priors). In Stage 2, the party leader decides (i) a conditional offer of pork to backbenchers, $r \in [0, R]$, and (ii) whether to make the vote a whip vote or not. In Stage 3, legislators vote between the alternatives x and q . A strategy for the leader is therefore a choice of a couple (a, r) , where $a \in \{m, n\}$ and $r \in [0, R]$. A strategy for a backbencher i is a pair of functions $\mathcal{Z}_i^m(\cdot; r)$ and $\mathcal{Z}_i^n(\cdot; r)$ mapping the set of types Θ and possible pork offers to party members $[0, R]$ to a vote in $\{q, x\}$ in the whip and no-whip votes, respectively. An equilibrium is a strategy profile

see for example Diermeier and Myerson (1999), Groseclose and Snyder (1996), Lizzeri and Persico (2001), Myerson (1993), and Snyder (1991).

⁸ The fact that $G(\cdot)$ is public information is irrelevant; the relevant assumption here is that there are no leadership challenges in the opposition, which we assume for simplicity of exposition.

$((a, r), \{\mathcal{Z}_i^p(\cdot; r), \mathcal{Z}_i^n(\cdot; r)\}_i)$ such that (i) (a, r) is feasible and sequentially rational and that (ii) $\{\mathcal{Z}_i^p(\cdot; r)\}_i$ and $\{\mathcal{Z}_i^n(\cdot; r)\}_i$ constitute, respectively, a Bayesian Nash Equilibrium of the whip and no-whip voting games.

PARTY DISCIPLINE: A DEFINITION

Defining party discipline involves two key considerations. First, party discipline should not reflect unity in voting that is driven by the absence of conflict between backbenchers over their preferred alternative, but instead should provide a measure of the power of the party leadership to induce cohesive or conformist behavior of backbenchers even in spite of significant internal dissent” (Krehbiel 1993, Cox and McCubbins 1993, Tsebelis 1995). A useful definition of party discipline must therefore capture the difference in backbenchers’ voting behavior between a non-partisan and a partisan framework. The second key consideration is the definition of the non-partisan benchmark to employ. In the model, I distinguish between payments that can be distributed on the spot to both members of the party and the opposition (*pork*), and promises of future partisan benefits, such as nomination to party lists. Should pork and promises of partisan benefits be treated symmetrically in the definition of party discipline? I put forward here the argument that a measure of discipline should not include changes in backbenchers’ voting behavior that are achieved with resources that could have otherwise been destined to non-party members (i.e., pork), allocating pork to party members means having to *buy* their support, and is therefore not an indication of power within the organization, but to the contrary one of weakness. Thus in this context, discipline is a measure of the power of partisan promises to induce behavior different than in a non-partisan benchmark. I implement the concept of discipline by considering the difference in the *equilibrium cutpoints* (the ideal policy of the most left-winged backbencher supporting the leader’s party line in equilibrium) in whip and no-whip votes.

Definition 1 Define party discipline, $d : [0, R] \rightarrow \mathbb{R}$, by

$$d(r) \equiv \inf\{\theta_i : \mathcal{Z}_i^n(\theta_i; r) = x\} - \inf\{\theta_i : \mathcal{Z}_i^p(\theta_i; r) = x\}$$

The comparison of cutpoints allows us to disentangle unity in voting that is due to the absence of conflict from unity in voting that is due to the power of the partisan promises, without being (directly) influenced by the distribution of preferences within the party (e.g., heterogeneity of backbenchers’ preferences, σ). This measure of discipline, moreover, is easily implementable in empirical analysis of voting in legislatures, and largely consistent with those employed in recent empirical studies of voting in the US Congress (see in particular Snyder and Groseclose (2000), McCarty *et al.* (2001), and Cox and Poole (2002)).⁹

⁹ An alternative measure of discipline achieving our two desiderata is to consider the fraction of the party voting for the party line in whip and no-whip votes, provided that we rule out the direct

EQUILIBRIA IN WHIP AND NO-WHIP VOTES

In this section, I characterize equilibria in whip votes and no-whip votes, and show that under some conditions equilibria of both games can be completely characterized by unique cutpoints δ_w and δ_n , so that discipline $d = \delta_n - \delta_w$ is uniquely defined as well. Consider first no-whip votes. Note that the net payoff of voting for x for a backbencher i is here given by $\Pi_n(\theta_i) = r - v(\theta_i)$, and is therefore independent of the actions of other players (this is a decision problem). Letting $\delta_n(r) \equiv v^{-1}(r)$, we then have

Remark 1 (No-Whip Votes) *In a no-whip vote, $\mathcal{X}_i^n(\theta_i; r) = x$ for all i such that $\theta_i > \delta_n(r)$ and $\mathcal{X}_i^n(\theta_i; r) = q$ for all i such that $\theta_i < \delta_n(r)$.*

The situation is qualitatively different in a whip vote. In a whip vote, only backbenchers with sufficiently *extreme* position-taking preferences (relative to the value of the resources at the leader's disposal; see Remark 3) are impervious to the actions of fellow party members. The decision of *moderate* backbenchers, instead, is determined by their beliefs about what others will do. For these individuals, supporting the leader's party line is optimal only if doing so allows them to capture a sufficiently high level of expected party payments. The net expected value of the leader's offer for individual i depends, in turn, on whether the leader will be able to retain the command of the party, and thus on i 's beliefs about the proportion of backbenchers supporting the leader's party line. If i believes that more than μ backbenchers will stick with the leader, he will want to do so as well; if he believes that at least $1 - \mu$ backbenchers will defect, he will defect too. In particular, if the distribution of party members' preferences is *common knowledge*, and the proportion of extremists is not high enough to determine the outcome of the leader's survival from the outset, radically different behavioral patterns can be sustained as equilibria by self-fulfilling beliefs (see Remark 3 in the Appendix). The assumption that the distribution of party members' preferences is common knowledge among backbenchers, however, is not desirable *per se*. Moreover, as recent developments in the *global games* literature show, relaxing this assumption allows us to pin down a unique equilibrium (Morris and Shin 1998, 2003, Frankel *et al.* 2001).

Consider a symmetric strategy profile in which backbenchers employ switching strategies with an arbitrary cutpoint δ . Denote by $\Pi(\theta_i; \delta)$ the net expected benefit of supporting x for a backbencher with ideal policy θ_i given this strategy profile. Similarly, denote by $\Gamma(\theta, \delta)$ the proportion of backbenchers voting for x according to this strategy profile given a particular realization of θ . Since $\theta_i|\theta \sim N(\theta, \sigma^2)$, then $\Gamma(\theta, \delta) = 1 - \Phi(\frac{\delta - \theta}{\sigma})$, where $\Phi(\cdot)$ is the cdf of the standard normal. Hence $\Gamma(\theta, \delta) = 1 - \Phi(\frac{\delta - \theta}{\sigma}) < \mu \Leftrightarrow \theta < \delta - \sigma\Phi^{-1}(1 - \mu)$, so that

$$\Pi(\theta_i; \delta) = r + e[1 - 2 \Pr(\theta < \delta - \sigma\Phi^{-1}(1 - \mu)|\theta_i)] - v(\theta_i).$$

effect of changes in the distribution of ideal points within the party (such as the party median, or the heterogeneity of backbenchers' ideal points). I show in the paper that the results in fact do not change when we use this alternative definition.

By Bayes' Law, $\theta|\theta_i \sim N(\widehat{\theta}(\theta_i), \widehat{\eta}^2)$, where $\widehat{\theta}(\theta_i) \equiv \frac{\sigma^2\theta_0 + \eta^2\theta_i}{\sigma^2 + \eta^2}$ and $\widehat{\eta} \equiv \frac{\sigma\eta}{\sqrt{\sigma^2 + \eta^2}}$. Then define the function

$$P(\delta, \theta_i) \equiv 1 - 2\Phi\left(\frac{\delta - \sigma\Phi^{-1}(1 - \mu) - \widehat{\theta}(\theta_i)}{\widehat{\eta}}\right). \quad (1)$$

Intuitively, $P(\delta, \theta_i)$ is the net expected value of a promise of one dollar of future partisan benefits made conditional on supporting the leader's party line for an individual with ideal policy θ_i , when every backbencher uses a switching strategy with cutoff point δ . Then $\Pi(\theta_i; \delta) = r + eP(\delta, \theta_i) - v(\theta_i)$. Denoting by $p(\delta) \equiv P(\delta, \delta)$ the net expected value of a promise of one dollar of future partisan benefits for the *critical backbencher* with ideal policy δ , and letting $\pi(\delta) \equiv \Pi(\delta; \delta)$, we have

$$\pi(\delta) = r + ep(\delta) - v(\delta).$$

Lemma 3 in the appendix shows that (i) $p(\cdot)$ is a decreasing function, and that (ii) $|p'(\cdot)|$ is bounded above by a decreasing function of η which goes to zero as $\eta \rightarrow \infty$. Since the slope of $v(\cdot)$ is bounded away from zero, this implies that for sufficiently high η , $\pi(\cdot)$ is an increasing function and $\pi(\delta) = 0$ at exactly one point. Proposition 1 is then a rather straightforward application of similar results in the *global games* literature (see for example Morris and Shin (2003)).

Proposition 1 *Let $\delta_w \in \{\delta : \pi(\delta) = 0\} \neq \emptyset$. There exists a symmetric equilibrium of the whip vote game in which $\mathcal{Z}_i^w(\theta_i, r) = x$ for all i such that $\theta_i \geq \delta_w$ and $\mathcal{Z}_i^w(\theta_i, r) = q$ for all i such that $\theta_i < \delta_w$. Moreover, there exists a $\bar{\eta}$ such that whenever $\eta > \bar{\eta}$, $\{\delta : \pi(\delta) = 0\}$ has a single element, δ_w , and this equilibrium is unique.*

In the next section, I turn to the substantive analysis leading to the main conclusions of the paper. In doing so, assume throughout that the condition in Proposition 1 is met.

MAIN RESULTS

I begin by considering the power of a leader in a *majoritarian* party ($\mu = 1/2$). The first goal is to isolate the value of promises of future benefits. To do so, assume initially that the leader has no access to pork resources that can be distributed on the spot ($R = 0$). I show that in this setting, credible promises of future partisan benefits confer only limited strength to the party leader, and a result similar in spirit to Aldrich and Rohde's conditional party government (Aldrich and Rohde 1997) emerges: the leader will use promises of partisan benefits to support the party line only if the leader's incentives are aligned (*ex ante*) with those of the majority of the party. Recall that θ_0 denotes the ideal policy of the *ex ante* party median. Then, we have:

Proposition 2 *Consider the leader of a majoritarian party ($\mu = 1/2$), endowed with no pork resources to distribute on the spot ($R = 0$). Then (i) whip votes occur in equilibrium if*

and only if, *ex ante*, a majority of the party prefers x to q (i.e., $v(\theta_0) < 0$), and (ii) in whip votes, the *ex ante* median supports the party line: $\delta_m \leq \theta_0$.

Note that this result holds independently of the level of partisan benefits available to the leader. The general moral is that even if credible *per se* and significant in amount, promises of future benefits will be relevant to alter voting behavior only if party members believe that the leader has a strong hold to the reins of power. Proposition 2 shows that with no resources to distribute on the spot, the leadership of a majoritarian institution is instead unstable (or more appropriately conditionally stable) and influence follows from the bottom up: promises are useful (and used) only if *ex ante* a majority of the party agrees with the leader on the ranking of alternatives in the first place.

To see the intuition for this result, consider Figure 1. Recall that backbenchers use the information contained in their own ideal policy, together with the location of the *ex ante* median to estimate the *ex post* distribution of preferences within the party. A backbencher with ideal policy equal to the *ex ante* median believes he/she is exactly centrist within the party, with half of the party being more right-winged and half of the party being more left-winged than himself. If this backbencher is also the critical backbencher, he

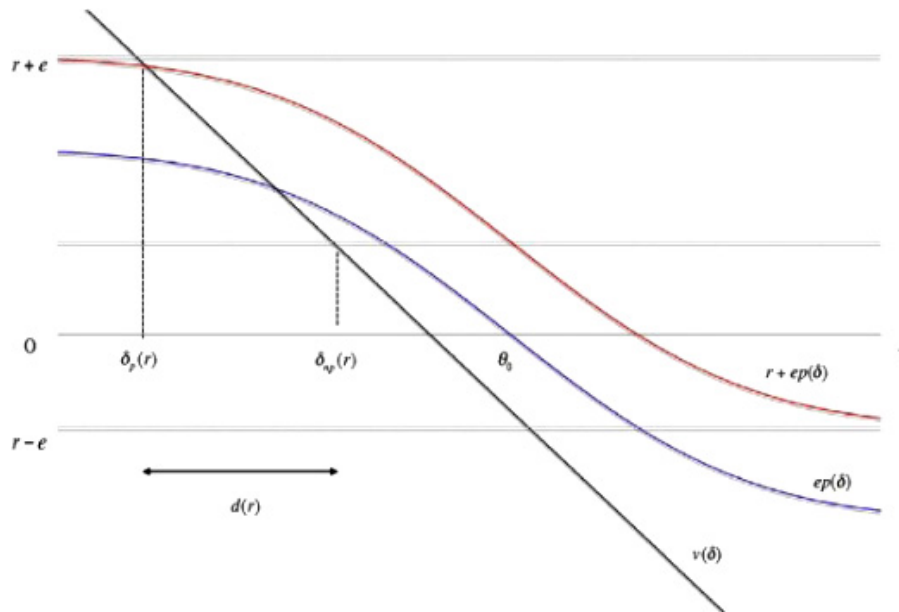


Figure 1. Equilibrium in whip and no-whip votes and party discipline.

Note: The functions $ep(\delta)$ and $v(\delta)$ plot the value of promises of future benefits, and the net position-taking cost of supporting the party line, for a critical backbencher with ideal point δ . The equilibrium cutpoint in whip votes is the point at which $r + ep(\delta) = v(\delta)$. Party discipline is the difference between the cutpoints in whip and no-whip votes.

also believes that exactly half of the party will support the party line. But then this critical *ex ante* party-centrist backbencher must attach probability 1/2 to the leader falling, and therefore a (net) value of zero to his/her promises (note in the figure that $p(\delta) = 0$ at $\delta = \theta_0$). Since the net expected value of the leader's promises of partisan benefits $ep(\delta)$ is continuously decreasing in δ , but everywhere flatter than $v(\delta)$, then for positive discipline we must have $\delta_p < \theta_0$. But this is only possible if the *ex ante* median θ_0 prefers x to q (if $\theta_0 > \delta_{np} = \frac{x+q}{2}$).

The next proposition shows that the (limited) power that electoral benefits confer to the leader in this environment can be attributed entirely to the heterogeneity of policy preferences among party backbenchers. That is, party discipline is monotonically increasing with the heterogeneity of party members' preferences, and vanishes in the limit as this heterogeneity goes to zero.

Proposition 3¹⁰ *Consider the leader of a majoritarian party ($\mu = 1/2$), endowed with no pork resources to distribute on the spot ($R = 0$). In equilibrium, party discipline decreases with the homogeneity of backbenchers' preferences, and $\lim_{\sigma \rightarrow 0} d = 0$.*

To get an intuition for this result, recall that backbenchers use both (i) public information about the central tendency of the party, and (ii) the information contained in their own preferences to form beliefs about the distribution of fellow party members' preferences (and thus ultimately about their actions). When party members' preferences are heterogeneous, only the *ex ante* median believes that he is centrist within the party, attaching equal probability to any member having ideal policy higher or lower than his/her own. Backbenchers with ideal policy $\theta_i < \theta_0$, instead, believe that a majority of party members have ideal policies that are to the left of the *ex ante* median. The informational content of a backbencher's ideal policy, in turn, increases with the homogeneity of the party. This implies, in particular, that backbenchers with ideal policy $\theta_i < \theta_0$ will attach a higher probability to the leader being overthrown (and thus a lower value to his/her promises of partisan benefits) the more homogeneous the party is. Note, however, that we are not concerned with how any arbitrary backbencher forms its beliefs, but with how the *critical* backbencher δ_m does. But we know from Proposition 2 that in majoritarian parties, the *ex ante* median must support the party line in whip votes, so that $\delta_m < \theta_0$. Therefore if discipline is positive, it must decrease with an increase in the homogeneity of backbenchers' preferences.

Heterogeneity is crucial here in that it allows the possibility of conflict, which the leader can exploit to bolster the support for the party line. In fact, as the previous result shows, in the bare setting of Proposition 3 this is all the leader has, and as backbenchers are almost perfectly homogeneous, a leader that needs the support of a majority of the party is powerless against the collective principal. In other words, it is the only when there is conflict and uncertainty about the position of fellow party legislators that the leader can be powerful with (just) promises of future benefits. While this result might at first appear counterintuitive in the light of Aldrich and Rohde's conditional party government

¹⁰ For the proof, see that of Proposition 5, which includes this as a special case.

hypothesis, the difference is entirely due to the costless and perfect coordination that is assumed in their framework, *vis-à-vis* the imperfect coordination that is allowed in this paper. In their work, as in this, increased heterogeneity benefits the leader (due to agenda setting power in the case of Aldrich and Rohde, and due to the lack of common knowledge of play, and its impact on the resolution of the collective action problem among backbenchers here). But while here the leader actually benefits from his/her increased power, in the world of Aldrich and Rohde backbenchers would simply remove powers from the leader under these circumstances.

It should be noted that this result does not depend on the fact that our definition of discipline considers only the difference in cutpoints between the whip vote and no-whip vote benchmarks. Suppose instead that we defined an alternative measure of party discipline, D , as the difference in the proportion of backbenchers voting for x in the whip and no-whip votes. Since $\theta_i|\theta \sim \mathcal{N}(\theta, \sigma^2)$, for any given θ this is

$$D(\delta_n(\sigma), \delta_m(\sigma); \sigma, \theta) = \Phi\left(\frac{\delta_n(\sigma) - \theta}{\sigma}\right) - \Phi\left(\frac{\delta_m(\sigma) - \theta}{\sigma}\right).$$

The total change in D brought by an infinitesimal change in heterogeneity σ is

$$\frac{dD}{d\sigma} = \frac{\partial D}{\partial \delta_n} \frac{\partial \delta_n}{\partial \sigma} + \frac{\partial D}{\partial \delta_m} \frac{\partial \delta_m}{\partial \sigma} + \frac{\partial D}{\partial \sigma}.$$

To isolate the effect of σ on discipline that is due to behavior, we want to rule out the direct effect of homogeneity on discipline. Thus, excluding the direct effect $\frac{\partial D}{\partial \sigma}$, we have

$$\frac{dD}{d\sigma} \Big|_{\frac{\partial D}{\partial \sigma}=0} = \frac{1}{\sigma} \left[\phi\left(\frac{\delta_n(\sigma) - \theta}{\sigma}\right) \frac{\partial \delta_n}{\partial \sigma} - \phi\left(\frac{\delta_m(\sigma) - \theta}{\sigma}\right) \frac{\partial \delta_m}{\partial \sigma} \right] = -\frac{1}{\sigma} \phi\left(\frac{\delta_m(\sigma) - \theta}{\sigma}\right) \frac{\partial \delta_m}{\partial \sigma},$$

since $\frac{\partial \delta_n}{\partial \sigma} = 0$. Therefore $\frac{dD}{d\sigma} \Big|_{\frac{\partial D}{\partial \sigma}=0} \leq 0$ if and only if $\frac{\partial \delta_m}{\partial \sigma} \geq 0$.

To summarize the results so far, I have shown that (i) if the leader of a majoritarian party is not endowed with pork resources to distribute on the spot, partisan benefits are worthless to oppose the (*ex ante*) majority of the party, and that (ii) if instead the leader sides with the *ex ante* majority of the party, the power of partisan resources is increasing in the heterogeneity of preferences within the party. Do these results change if the leader can distribute pork? Interestingly, the answers are yes, and no, respectively.

Consider first the impact of pork on discipline and our *sui generis* conditional party government. While in the absence of pork, promises of partisan benefits have no value to oppose a majority of the party, with (enough) pork this is no longer true, and whip votes can exist in equilibrium even if $v(\theta_0) > 0$. In this case, however, the allocation of pork to party members has to be at least as large as to attain the support of the (*ex ante*) party median, i.e., $r \geq -v(\theta_0)$. This result is due to a *complementarity* between the allocation of pork to party members and the value of promises of future partisan benefits. In a no-whip vote — where backbenchers' beliefs about the actions of fellow party members are irrelevant — decreasing the allocation of pork to the party by one dollar leads to an

equivalent reduction in the value of the leader's offer. In a whip vote, instead, the value of the leader's promises of partisan benefits is tied to the fate of the leader. But a reduction in the allocation of pork to party members will induce backbenchers to anticipate a lower aggregate support for the party line and, as a result, a higher probability of the leader being overthrown, leading ultimately to a depreciation of the value of the leader's promises of partisan benefits. The same logic implies, in fact, that

Remark 2 *Increasing the offers of pork to backbenchers at the expense of opposition legislators increases party discipline, i.e., discipline $d(r) = \delta_n(r) - \delta_w(r)$ is increasing in r .*

Due to the complementarity between pork and the value of promises of partisan benefits, increasing the offers of pork to backbenchers increases party discipline. This implies in turn that using pork to buy the support of legislators in the opposition has an opportunity cost, as buying the opposition means weakening the support inside the party. The next result builds on the fact that this opportunity cost of buying the support of the opposition is larger the more easily contestable the leadership is (the larger μ is). Note first that party discipline is decreasing in μ . Indeed, for every $\mu \in (0, 1/2]$ there is a minimum level $r_{\min}(\mu)$ that offers to backbenchers must surpass in order for a whip vote to be profitable for the leader, and it can be easily verified that $r_{\min}(\mu)$ is an increasing function, with maximum at $r_{\min}(1/2) = -v(\theta_0)$. The next proposition shows, moreover, that when whip votes occur in equilibrium, the leader allocates less pork to buy opposition legislators the more contestable the leadership position is. This result holds even when the leader does not care about the leadership position *per se*. In essence, the result is due to the fact that increasing the contestability of the leadership boosts the complementarity between pork and the value of promises of partisan benefits. In this situation, *weak* leaders find more profitable buying their own party, thus avoiding large depreciations of the value of the future benefits at their disposal. Let $r^*(\mu)$ denote the optimal allocation of pork to the party in a whip vote given μ . Then we have

Proposition 4 *Suppose that given $\mu = \mu^0 \in (0, 1/2)$, it is optimal for the leader to induce a whip vote in equilibrium. Then it is also optimal for the leader to induce a whip vote in equilibrium if $\mu = \mu^1 < \mu^0$, and $r^*(\mu^1) \leq r^*(\mu^0)$. Moreover, if under μ^0 the leader allocates pork to both members of the party and the opposition, or $r^*(\mu^0) \in (r_{\min}(\mu^1), R)$, then the inequality is strict.*

In Proposition 3, it is showed that when the leader of a majoritarian party ($\mu = 1/2$) is not endowed with pork resources to distribute on the spot ($R = 0$), party discipline decreases with the homogeneity of backbenchers' preferences, and that in fact $\lim_{\sigma \rightarrow 0} d = 0$. Proposition 5 revisits this result, allowing for arbitrary majority requirements for removal and allocations of pork to party members. I show provided that $\mu = 1/2$, the result does generalize to arbitrary allocations of pork to backbenchers $r \leq R$ as stated. When $\mu < 1/2$, instead, the main intuition described above breaks down, and this is no longer the case. The gist of the argument is that with $\mu < 1/2$, it is possible for the *ex ante* party median to oppose the party line, while still having positive

discipline. I also show that in this case $\text{Lim}_{\sigma \rightarrow 0} d > 0$, so that the leader endowed with promises of partisan benefits is powerful even against a perfectly homogeneous party.

Proposition 5 *Let μ and $r \in [0, R]$ be given. If the party is majoritarian ($\mu = 1/2$), then in equilibrium discipline decreases with the homogeneity of backbenchers' preferences, and $\text{Lim}_{\sigma \rightarrow 0} d = 0$. If instead $\mu < 1/2$, then (i) discipline does not necessarily decrease with homogeneity, and (ii) $d^* \equiv \text{Lim}_{\sigma \rightarrow 0} d$ is positive and increasing in e and $1 - \mu$.*

This result should not be interpreted as implying that discipline will be decreasing in the heterogeneity of the party when $\mu < 1/2$, but only that in this case the comparative static result will not hold uniformly. In fact, a sufficient condition for discipline to decrease with homogeneity is that the (*ex ante*) median supports the party line in equilibrium, which always holds in a whip vote in the case of a majoritarian party.

CONCLUDING REMARKS

This paper builds on two basic elements characterizing leadership in political parties. On the one hand, party leaders rarely have unconditional control of party resources. Instead, their authority relies on an implicit contract with backbenchers that grants them legitimacy: leaders manage party resources conditional on the support of the organization. On the other hand, the *party* is at best a collective, heterogeneous, principal. Under these circumstances, the leader does not merely solve problems of coordination, but instead *de facto* resolves conflict, leading members to pursue policies with less than total support within the party. To capture these forces within a unifying framework, I propose the concept of contestable leadership. In a contestable leadership, the direction of the party line can be disputed within the party, but doing so requires that a sufficiently large number of heterogeneous backbenchers successfully coordinate their opposition to the party leadership. I argue that the key element delimiting the power of the leader is how this impure coordination problem among backbenchers is resolved.

I show that rewards that can be distributed publicly and on the spot are effective tools to coordinate beliefs about the stability of the leader, and thus also to increase the value of the leader's promises of future benefits (and with it party discipline). These spot resources are in fact necessary for the leader to be powerful: without them, the leader can use promises of future benefits to sway members' behavior only if a majority of the party agrees (*ex ante*) with the leader's preferred position in the first place. In a nutshell, discipline requires that challenging the party line is perceived to be difficult among backbenchers, and something other than promises is needed to pin down these beliefs. A similar result applies in fact as well with respect to the party line, assuming that the party leader has agenda setting power: an *extreme* party line (relative to the preferred position of the *ex ante* median) conveys information to all backbenchers about a general discontent with the leader's choice, and thus lowers the impact of promises of future benefits. An obvious implication of this logic is that discipline falls with x faster than in a unidirectional model of influence, and faster the more vulnerable the leader is

to internal challenges, but a more complete analysis of this topic is beyond the scope of this paper. Other avenues for future research also seem worth pursuing. A natural next step is to relax the simplicity of the current vote buying model to study optimal size and composition of partisan coalitions. Modeling dynamics would also allow us to analyze whether a leader would optimally engage in vote buying at the initial phase or the end of his/her tenure, and to focus on the dynamics of discipline and revolts (Angeles *et al.* 2007).

APPENDIX

Remark 3 Let $\underline{\theta}_i \equiv v^{-1}(r+e)$ and $\bar{\theta}_i \equiv v^{-1}(r-e)$. Suppose that θ is common knowledge, and that $\theta_i < \theta + \sigma\Phi^{-1}(1-\mu) < \bar{\theta}_i$.¹¹ Then the following strategy profiles are BNE of the party whip voting game:

- (1) $\mathcal{X}_i^p(\theta_i; r) = x \forall i : \theta_i > \underline{\theta}_i$ and $\mathcal{X}_i^p(\theta_i; r) = q \forall i : \theta_i < \underline{\theta}_i$ and
- (2) $\mathcal{X}_i^p(\theta_i; r) = x \forall i : \theta_i > \bar{\theta}_i$ and $\mathcal{X}_i^p(\theta_i; r) = q \forall i : \theta_i < \bar{\theta}_i$.

Proof: Consider first strategy profile 1. Since $\theta_i|\theta \sim N(\theta, \sigma^2)$, the proportion of backbenchers voting for x is then given by $1 - \Phi\left(\frac{\theta_i - \theta}{\sigma}\right)$, where $\Phi(\cdot)$ is the cdf of the standard normal. The leader survives the challenge (with certainty) if

$$1 - \Phi\left(\frac{\theta_i - \theta}{\sigma}\right) > \mu \Leftrightarrow \theta > \underline{\theta}_i - \sigma\Phi^{-1}(1 - \mu).$$

Since this is true by hypothesis, the expected net payoff of voting for q for backbencher i is given by $v(\theta_i) - r - e$. Then optimality implies $\mathcal{X}_i^p(\theta_i; r) = q$ if $\theta_i < v^{-1}(r+e) \equiv \underline{\theta}_i$ and $\mathcal{X}_i^p(\theta_i; r) = x$ if $\theta_i > \underline{\theta}_i$. Similarly, consider strategy profile 2. The proportion of backbenchers voting for x is then given by $1 - \Phi\left(\frac{\bar{\theta}_i - \theta}{\sigma}\right)$. The leader will fall for sure if $1 - \Phi\left(\frac{\bar{\theta}_i - \theta}{\sigma}\right) < \mu \Leftrightarrow \theta < \bar{\theta}_i - \sigma\Phi^{-1}(1 - \mu)$, which again is true by hypothesis. The expected net payoff of voting for x for backbencher i is then given by $r - e - v(\theta_i)$, and optimality implies $\mathcal{X}_i^p(\theta_i; r) = x \forall i : \theta_i > \bar{\theta}_i$ and $\mathcal{X}_i^p(\theta_i; r) = q \forall i : \theta_i < \bar{\theta}_i$. ■

Proof of Proposition 1

The following definitions are used here. For a given strategy profile of the whip vote game $\{\mathcal{X}_i^p\}$, where each $\mathcal{X}_i^p : \Theta \times [0, R] \rightarrow \{q, x\}$, let $\xi(z)$ denote the proportion of backbenchers for whom $\mathcal{X}_i^p(z; r) = x$, let $\Gamma(\theta; \xi)$ denote the proportion of backbenchers that would end up supporting x given a particular realization of θ and an aggregate voting mapping ξ , and let $\Pi(\theta_i; \xi)$ denote the expected net benefit of supporting x for a backbencher with ideal policy θ_i , given ξ . Proposition 1 follows from three lemmas. In

¹¹ When $\theta + \sigma\Phi^{-1}(1 - \mu) < \underline{\theta}_i$, strategy profile 1 in the remark constitutes the unique BNE of the party whip vote game. Similarly, when $\theta + \sigma\Phi^{-1}(1 - \mu) > \bar{\theta}_i$, strategy profile 2 is the unique BNE.

Lemma 1, I show that (i) $\{\delta : \pi(\delta) = 0\} \neq \emptyset$, and that (ii) with $\delta_m \in \{\delta : \pi(\delta) = 0\}$, there exists a symmetric equilibrium of the whip vote game in which $\mathcal{X}_i^m(\theta_i, r) = x$ for all i such that $\theta_i \geq \delta_m$ and $\mathcal{X}_i^m(\theta_i, r) = q$ for all i such that $\theta_i < \delta_m$. In Lemma 2, I show that if $\pi(\delta)$ is strictly increasing $\{\delta : \pi(\delta) = 0\}$ has a single element δ_m , and this equilibrium is unique. The next step is thus to provide a sufficient condition for $\pi(\delta)$ to be strictly increasing. Note that this happens iff $ep'(\delta) > v'(\delta)$ for every δ , and that we know already that $v(\cdot)$ is a strictly decreasing function. Lemma 3 shows that while $p(\delta)$ is also a decreasing function, it can be made arbitrarily flat by reducing the precision of public information (by increasing η). Specifically, for any $Q > 0$, there exists a $\bar{\eta}(Q)$ such that if $\eta > \bar{\eta}(Q)$, then $|p'(\delta)| < Q$. Then $\pi(\delta)$ is strictly increasing if $\eta > \bar{\eta}(\frac{1}{\epsilon}|v'(\delta)|)$, and we are done. ■

Lemma 1 $\{\delta : \pi(\delta) = 0\} \neq \emptyset$. Let $\delta_m \in \{\delta : \pi(\delta) = 0\}$. There exists a symmetric equilibrium of the whip vote game in which $\mathcal{X}_i^m(\theta_i, r) = x$ for all i such that $\theta_i \geq \delta_m$ and $\mathcal{X}_i^m(\theta_i, r) = q$ for all i such that $\theta_i < \delta_m$.

Proof: Our first task is to show that $\{\delta : \pi(\delta) = 0\} \neq \emptyset$. Consider the points $\underline{\theta}_i \equiv v^{-1}(r + e)$ and $\bar{\theta}_i \equiv v^{-1}(r - e)$ that were defined in Remark 3. Note that the net payoff of voting for q for backbencher i in the event that the leader survives the challenge is given by $v(\theta_i) - r - e$. Since the net payoff of voting for q for backbencher i is always at least $v(\theta_i) - r - e$, then $\theta_i < \underline{\theta}_i \Rightarrow \Pi(\theta_i; \xi) < 0$ for any ξ . Similarly, since the net payoff of voting for x for backbencher i is always at least $r - e - v(\theta_i)$ (where the challenge is successful for sure), then $\theta_i > \bar{\theta}_i \Rightarrow \Pi(\theta_i; \xi) > 0$ for any ξ . It should be noted that the points $\bar{\theta}_i$ and $\underline{\theta}_i$ are well defined, since $v(\cdot)$ is continuous, decreasing, and $\text{Lim}_{\theta_i \rightarrow -\infty} v(\theta_i) = \infty$, while $\text{Lim}_{\theta_i \rightarrow \infty} v(\theta_i) = -\infty$. Now, $\pi(\delta) \equiv \Pi(\delta, \delta) \equiv \Pi(\theta_i = \delta; \xi = 1_{\{\theta_i \geq \delta\}})$. Then the previous argument implies, in particular, that $\pi(\delta) > 0$ for $\delta > \bar{\theta}_i$, and $\pi(\delta) < 0$ for $\delta < \underline{\theta}_i$. Since $\pi(\delta)$ is continuous, this implies that $\{\delta : \pi(\delta) = 0\} \neq \emptyset$. Next, let $\delta_m \in \{\delta : \pi(\delta) = 0\}$. To show the existence of the symmetric equilibrium, it is now enough to show that $\Pi(\theta_i; 1_{\{\theta_i \geq \delta\}})$ is increasing in θ_i . But it is easy to see from 1 that $P(\delta, \theta_i)$ is increasing in θ_i . Since $v(\theta_i)$ is decreasing, the result follows. ■

Lemma 2 Suppose that $\pi(\delta)$ is strictly increasing. Then $\{\delta : \pi(\delta) = 0\}$ has a single element δ_m , and the equilibrium of Lemma 1 is unique.

Proof (Morris and Shin (1998)) If $\pi(\delta)$ is strictly increasing, there is a unique δ_m solving $\pi(\delta) = 0$. I show next that this in turn implies that the symmetric equilibrium with switching strategies at δ_m is the unique equilibrium. So consider any equilibrium of the game, and define the numbers

$$\underline{z} \equiv \inf\{z | \xi(z) > 0\} \quad \text{and} \quad \bar{z} \equiv \sup\{z | \xi(z) < 1\}.$$

Note first that

$$\begin{aligned} \bar{z} &\equiv \sup\{z | \xi(z) < 1\} \geq \sup\{z | 0 < \xi(z) < 1\} \geq \inf\{z | 0 < \xi(z) < 1\} \\ &\geq \inf\{z | \xi(z) > 0\} \equiv \underline{z}. \end{aligned} \tag{A.1}$$

Now, for any $z \in \{z | \xi(z) > 0\}$, there is some i for which $x_i^p(z; r) = x$. This is only consistent with equilibrium behavior if the payoff to supporting x (for individual i and for anyone else, since they are all identical, *ex ante*) is at least as high as the payoff to supporting q given ideal policy z , i.e., $\Pi(z, \xi) \geq 0$. By continuity, this is also true at \underline{z} , i.e.,

$$\Pi(\underline{z}, \xi) \geq 0. \quad (\text{A.2})$$

Now consider the payoff $\Pi(\underline{z}, 1_{\{\theta_i \geq \underline{z}\}})$. It is clear that, for any z , $1_{\{\theta_i \geq \underline{z}\}}(z) \geq \xi(z)$. But — in general — whenever $\xi(z) \geq \xi'(z)$ for any z , then $\Pi(z, \xi) \geq \Pi(z, \xi')$. Hence $\Pi(\underline{z}, 1_{\{\theta_i \geq \underline{z}\}}) \geq \Pi(\underline{z}, \xi)$ for any \underline{z} , and in particular

$$\pi(\underline{z}) \equiv \Pi(\underline{z}, 1_{\{\theta_i \geq \underline{z}\}}) \geq \Pi(\underline{z}, \xi). \quad (\text{A.3})$$

Thus combining (A.2) and (A.3) I obtain

$$\pi(\underline{z}) \geq 0. \quad (\text{A.4})$$

Now by hypothesis, $\pi(\delta)$ is increasing in δ . Since δ_m is the unique value of δ which solves $\pi(\delta) = 0$, this means $\underline{z} \geq \delta_m$. A symmetric argument establishes that $\bar{z} \leq \delta_m$. Thus $\bar{z} \leq \delta_m \leq \underline{z}$. This together with (A.1) implies that $\underline{z} = \delta_m = \bar{z}$. Thus in any equilibrium the x 's aggregate support mapping ξ , and thus the strategy of every backbencher, x_i^p , is given by $1_{\{\theta_i \geq \delta_m\}}$. ■

Lemma 3 $p(\cdot)$ is a decreasing function of δ . Furthermore, for any $Q > 0$, there exists a $\bar{\eta}(Q)$ such that if $\eta > \bar{\eta}(Q)$, then $|p'(\delta)| < Q$.

Proof: Since

$$\left(\frac{\theta - \hat{\theta}(\theta_i)}{\hat{\eta}} \right)_{\theta = \delta - \sigma\Phi^{-1}(1-\mu)} = \frac{1}{\hat{\eta}} \left[\frac{\sigma^2(\delta - \theta_0) + \eta^2(\delta - \theta_i)}{\sigma^2 + \eta^2} - \sigma\Phi^{-1}(1 - \mu) \right],$$

then

$$p(\delta) = P(\delta, \delta) = 1 - 2\Phi \left(\frac{1}{\hat{\eta}} \left[\frac{\sigma^2}{\sigma^2 + \eta^2} (\delta - \theta_0) - \sigma\Phi^{-1}(1 - \mu) \right] \right).$$

Hence

$$\frac{\partial p(\delta)}{\partial \delta} = -2\phi \left(\frac{1}{\hat{\eta}} \left[\frac{\sigma^2}{\sigma^2 + \eta^2} (\delta - \theta_0) - \sigma\Phi^{-1}(1 - \mu) \right] \right) \frac{1}{\eta} \frac{1}{\sqrt{1 + \frac{\eta^2}{\sigma^2}}}.$$

That $p'(\delta) < 0$ follows immediately. And since $|p'(\delta)|$ is bounded above by $\frac{2}{\eta}$, $|p'(\delta)| < Q$ for $\eta > 2/Q = \bar{\eta}(Q)$. ■

Proof of Proposition 2

First note that the leader will call a whip vote in equilibrium if and only if discipline is positive. Now, $d = \delta_n - \delta_m \geq 0 \Leftrightarrow p(\delta_m) \geq 0$. That is, discipline is positive if and only if the critical backbencher δ_m assigns net positive value to the leader's promises of electoral benefits. But with $\mu = 1/2$, $p(\delta_m) \geq 0 \Leftrightarrow \delta_m \leq \theta_0$, because

$$\Pr(\Gamma(\theta, \delta_m) < \mu | \theta_i = \delta_m) = \Pr(\theta < \delta_m | \theta_i = \delta_m) < 1/2 \Leftrightarrow \delta_m < \theta_0.$$

That is, with $\mu = 1/2$, the critical backbencher δ_m assigns net positive value to the leader's promises of electoral benefits if and only if the *ex ante* party median is in the leader's coalition (iff $\delta_m < \theta_0$). Hence $d \geq 0 \Leftrightarrow \delta_m \leq \theta_0$. Now, with $r = 0$, $\delta_n = v^{-1}(0)$, and then $v(\delta_n) = 0$. Since $ep(\delta)$ is continuously decreasing, but everywhere flatter than $v(\delta)$, then $\delta_n \leq \theta_0 \Leftrightarrow \delta_m \leq \delta_n \Leftrightarrow d \geq 0$. Finally, $x \stackrel{\theta_0}{\geq} q \Leftrightarrow \delta_n = v^{-1}(0) \leq \theta_0$, implying that $x \stackrel{\theta_0}{\geq} q \Leftrightarrow d \geq 0$. ■

Proof of Remark 2

Note that discipline $d = \delta_n - \delta_m$ is increasing in r if and only if $-\partial\delta_n/\partial r < -\partial\delta_m/\partial r$. But since $\delta_n = v^{-1}(r)$, and δ_m is $r + ep(\delta_m) - v(\delta_m) \equiv 0$, then $-\partial\delta_n/\partial r = \frac{1}{-v'(\delta_n)} < \frac{1}{ep'(\delta_m) - v'(\delta_m)} = -\partial\delta_m/\partial r$, since $p'(\delta_m) < 0$, $\delta_m < \delta_n$, and $|v(\cdot)|$ is convex. ■

Proof of Proposition 4

The first step is to characterize optimal allocations of pork to party members under rule μ , $r^*(\mu)$. Let $H(\cdot) \equiv [1 - G(v^{-1}(\cdot))]$. The mass of legislators in the opposition voting for x given pork offer r_0 is given by $H(r_0)\beta$. Note that $H'(r_0) \geq 0$ for all r_0 . Pork resource constraint is given by $r_0\beta + r \leq R$. Since this will hold with equality in the optimum, I write $r_0 = \frac{R-r}{\beta}$. Denoting then the outcome as $y \in \{q, x\}$, then, $y = x$ conditional on a realization θ if and only if $H(\frac{R-r}{\beta})\beta + \Gamma(\theta, \delta_p(r, \mu)) \geq \frac{(1+\beta)}{2}$. Since $\Gamma(\theta, \delta_p(r, \mu)) = 1 - \Phi(\frac{\delta_p(r, \mu) - \theta}{\sigma})$, this is

$$\theta \geq \delta_m(r, \mu) - \mathcal{J}(r),$$

where $\mathcal{J}(r) \equiv \sigma\Phi^{-1}\left(\frac{(1-\beta)}{2} + H\left(\frac{R-r}{\beta}\right)\beta\right)$. Then for the leader,

$$\Pr(y = x) = 1 - \Phi\left(\frac{1}{\eta} [(\delta_m(r, \mu) - \theta_0) - \mathcal{J}(r)]\right).$$

An optimal allocation of pork for the leader $r^*(\mu)$ maximizes $\Pr(y = x)$. The FOC is

$$\left| \frac{\partial\delta_m(r^*(\mu), \mu)}{\partial r} \right| - \mathcal{J}'(r^*(\mu)) \begin{cases} > 0 & \text{and } r^*(\mu) = R \\ = 0 & \text{and } r^*(\mu) \in (r_{\min}(\mu), R) \\ < 0 & \text{and } r^*(\mu) = r_{\min}(\mu). \end{cases} \quad (\text{A.5})$$

The second and final step is to show that for all r ,

$$\left| \frac{\partial\delta_m(r, \mu^0)}{\partial r} \right| > \left| \frac{\partial\delta_m(r, \mu^1)}{\partial r} \right| \text{ whenever } \mu^0 > \mu^1, \quad (\text{A.6})$$

which implies that

$$\left| \frac{\partial \delta_w(r^*(\mu^1), \mu^0)}{\partial r} \right| > \left| \frac{\partial \delta_w(r^*(\mu^1), \mu^1)}{\partial r} \right| \text{ whenever } \mu^0 > \mu^1. \quad (\text{A.7})$$

Then (A.7) together with (A.5) imply that $r^*(\mu^0) \geq r^*(\mu^1)$. Moreover, if $r^*(\mu^1) \in (r_{\min}(\mu^1), R)$, so that $\left| \frac{\partial \delta_w(r^*(\mu^1), \mu^1)}{\partial r} \right| = \mathcal{F}'(r^*(\mu^1))$, then $\left| \frac{\partial \delta_w(r^*(\mu^1), \mu^0)}{\partial r} \right| > \mathcal{F}'(r^*(\mu^1))$, and hence $r^*(\mu^0) > r^*(\mu^1)$.

Note that for all r, μ ,

$$\left| \frac{\partial \delta_w(r, \mu)}{\partial r} \right|^{-1} = \left| \frac{\partial v(\delta_w(r, \mu))}{\partial \delta} \right| - e \left| \frac{\partial p(\delta_w(r, \mu); \mu)}{\partial \delta} \right|$$

so that (A.6) can be written as:

$$\begin{aligned} & e \left\{ \left| \frac{\partial p(\delta_w(r, \mu^0); \mu^0)}{\partial \delta} \right| - \left| \frac{\partial p(\delta_w(r, \mu^1); \mu^1)}{\partial \delta} \right| \right\} \\ & > \left| \frac{\partial v(\delta_w(r, \mu^0))}{\partial \delta} \right| - \left| \frac{\partial v(\delta_w(r, \mu^1))}{\partial \delta} \right|. \end{aligned} \quad (\text{A.8})$$

Note, next, that since in a whip vote $\delta_w(r, \mu)$ is increasing in μ , then $\delta_w(r, \mu^1) < \delta_w(r, \mu^0)$. The convexity of $|v(\cdot)|$ then implies that

$$|v'(\delta_w(r, \mu^1))| \geq |v'(\delta_w(r, \mu^0))|. \quad (\text{A.9})$$

Also, since

$$\left| \frac{\partial p(\delta; \mu)}{\partial \delta} \right| = 2\phi \left(\frac{1}{\tilde{\eta}} \left[\frac{\sigma^2}{\sigma^2 + \eta^2} (\delta - \theta_0) - \sigma \Phi^{-1}(1 - \mu) \right] \right) \frac{1}{\eta} \frac{1}{\sqrt{1 + \frac{\eta^2}{\sigma^2}}},$$

it can be verified that if $d > 0$ then $\frac{\partial}{\partial \mu} \left(\left| \frac{\partial p(\delta; \mu)}{\partial \delta} \right| \right) > 0$, so that

$$\left| \frac{\partial p(\delta_w(r, \mu^0); \mu^0)}{\partial \delta} \right| > \left| \frac{\partial p(\delta_w(r, \mu^0); \mu^1)}{\partial \delta} \right|, \quad (\text{A.10})$$

and that (ii) $\frac{\partial^2 p(\delta; \mu)}{\partial \delta^2} < 0$, so that $\delta_w(r, \mu^0) > \delta_w(r, \mu^1)$ implies that

$$\left| \frac{\partial p(\delta_w(r, \mu^0); \mu^1)}{\partial \delta} \right| > \left| \frac{\partial p(\delta_w(r, \mu^1); \mu^1)}{\partial \delta} \right|. \quad (\text{A.11})$$

Then (A.10) and (A.11) imply that

$$\left| \frac{\partial p(\delta_w(r, \mu^0); \mu^0)}{\partial \delta} \right| > \left| \frac{\partial p(\delta_w(r, \mu^1); \mu^1)}{\partial \delta} \right|. \quad (\text{A.12})$$

Therefore (A.9) and (A.12) imply that (A.8) holds. \blacksquare

Proof of Proposition 5

Note first that

$$p(\delta) = P(\delta, \delta) = 1 - 2\Phi\left(\frac{1}{\widehat{\eta}} \left[\frac{\sigma^2}{\sigma^2 + \eta^2} (\delta - \theta_0) - \sigma\Phi^{-1}(1 - \mu) \right]\right),$$

where

$$\frac{1}{\widehat{\eta}} \left[\frac{\sigma^2}{\sigma^2 + \eta^2} (\delta - \theta_0) - \sigma\Phi^{-1}(1 - \mu) \right] = \left(\frac{\theta - \widehat{\theta}(\theta_i = \delta)}{\widehat{\eta}} \right)_{\theta = \delta - \sigma\Phi^{-1}(1 - \mu)}.$$

Thus

$$\frac{\partial p(\delta; \sigma)}{\partial \sigma} = -2\phi(\cdot) \frac{1}{\sqrt{\sigma^2 + \eta^2}} \left[\frac{\eta}{\sigma^2 + \eta^2} (\delta - \theta_0) - \frac{\sigma}{\eta} \Phi^{-1}(1 - \mu) \right],$$

so that $\frac{\partial p(\delta; \sigma)}{\partial \sigma} \geq 0$ if and only if:

$$\theta_0 \geq \delta_w - \sigma\Phi^{-1}(1 - \mu) \left(1 + \frac{\sigma^2}{\eta^2} \right). \quad (\text{A.13})$$

But if $p(\delta; \sigma)$ increases with σ at $\delta_w(\sigma')$, then $\sigma'' > \sigma' \implies \delta_w(\sigma'') < \delta_w(\sigma')$. Hence more heterogeneity of backbenchers' preferences must in this case increase party discipline. Similarly, if $p(\delta; \sigma)$ decreases with σ at $\delta_w(\sigma')$, then more heterogeneity of backbenchers' preferences must in this case reduce discipline. Now,

$$d \geq 0 \Leftrightarrow p(\delta_w) \geq 0 \Leftrightarrow \left(\frac{\theta - \widehat{\theta}(\theta_i = \delta_w)}{\widehat{\eta}} \right)_{\theta = \delta_w - \sigma\Phi^{-1}(1 - \mu)} \leq 0.$$

That is, $d \geq 0$ if and only if

$$\theta_0 \geq \delta_w - \left(1 + \frac{\eta^2}{\sigma^2} \right) \sigma\Phi^{-1}(1 - \mu). \quad (\text{A.14})$$

Hence, in equilibrium, party discipline necessarily increases with σ if (A.13) is satisfied whenever (A.14) is. Since δ_w is a continuously decreasing function of θ_0 , bounded below by $\underline{\theta}_i \equiv v^{-1}(r + e)$ and above by $\overline{\theta}_i \equiv v^{-1}(r - e)$, there is a unique θ_0^* solving (A.13) with equality, and a unique θ_0^{**} solving (A.14) with equality. If $\mu = 1/2$, these two inequalities

collapse to $\theta_0 \geq \delta_w$. Therefore in equilibrium, discipline necessarily increases with σ . Moreover, $\delta_w = \theta_0 \Leftrightarrow p(\delta_w) = 0 \Leftrightarrow \theta_0 = v^{-1}(r)$, so that $\theta_0^* = \theta_0^{**} = v^{-1}(r)$. With $\mu < 1/2$, however, (A.13) is satisfied whenever (A.14) is only if $\sigma \geq \eta$.

To establish the results for the limit as $\sigma \rightarrow 0$, we show that

$$\lim_{\sigma \rightarrow 0} d = v^{-1}(r) - v^{-1}(r + e[1 - 2\mu]).$$

To see this, let

$$\begin{aligned} f(\sigma) &= \Phi \left(\left(\frac{\theta - \widehat{\theta}(\theta_i = \delta)}{\widehat{\eta}} \right)_{\theta = \delta - \sigma \Phi^{-1}(1 - \mu)} \right) \\ &= \Phi \left(\frac{1}{\eta} \left(1 + \frac{\eta^2}{\sigma^2} \right)^{-\frac{1}{2}} (\delta_w(\sigma) - \theta_0) - \left(1 + \frac{\sigma^2}{\eta^2} \right)^{\frac{1}{2}} \Phi^{-1}(1 - \mu) \right). \end{aligned}$$

Since $f(\sigma)$ is continuous in an interval around 0,

$$\lim_{\sigma \rightarrow 0} f(\sigma) = \Phi \left(\lim_{\sigma \rightarrow 0} \left[\frac{1}{\eta} \left(1 + \frac{\eta^2}{\sigma^2} \right)^{-\frac{1}{2}} (\delta_w(\sigma) - \theta_0) - \left(1 + \frac{\sigma^2}{\eta^2} \right)^{\frac{1}{2}} \Phi^{-1}(1 - \mu) \right] \right).$$

Note that $\lim_{\sigma \rightarrow 0} \left(1 + \frac{\sigma^2}{\eta^2} \right)^{\frac{1}{2}} = 1$, and $\lim_{\sigma \rightarrow 0} \frac{1}{\eta} \left(1 + \frac{\eta^2}{\sigma^2} \right)^{-\frac{1}{2}} = 0$. Since $\delta_w(\sigma)$ is bounded (by $\underline{\theta}_i$ and $\bar{\theta}_i$), this implies that

$$\lim_{\sigma \rightarrow 0} f(\sigma) = \Phi(-\Phi^{-1}(1 - \mu)) = \Phi(\Phi^{-1}(\mu)) = \mu.$$

Now,

$$v(\delta_w) \equiv r + e \left[1 - 2\Phi \left(\left(\frac{\theta - \widehat{\theta}(\theta_i = \delta)}{\widehat{\eta}} \right)_{\theta = \delta - \sigma \Phi^{-1}(1 - \mu)} \right) \right].$$

Therefore in the limit as $\sigma \rightarrow 0$, $v(\tilde{\delta}_w) = r + e[1 - 2\mu]$, so that

$$\lim_{\sigma \rightarrow 0} d = v^{-1}(r) - v^{-1}(r + e[1 - 2\mu]). \quad \blacksquare$$

REFERENCES

- Aldrich, J. H. 1995. *Why Parties? The Origin and Transformation of Party Politics in America*. Chicago, IL: University of Chicago Press.
- Aldrich, J. H., and D. W. Rohde. 1997. "The Transition to Republican Rule in the House: Implications for Theories of Congressional Politics." *Political Science Quarterly* 112: 541–567.

- Angeletos, G.-M., C. Hellwig, and A. Pavan. 2007. "Dynamic Global Games of Regime Change: Learning, Multiplicity, and Timing of Attacks." *Econometrica* 75: 711–756.
- Bowler, S., D. M. Farrell, and R. S. Katz. 1999a. *Party Cohesion, Party Discipline and Parliaments*.
- Bowler, S., D. M. Farrell, and R. S. Katz (eds.). 1999b. *Party Discipline and Parliamentary Government*. Ohio State University Press.
- Bueno de Mesquita, B., J. D. Morrow, A. Smith, and R. M. Siverson. 2003. *The Logic of Political Survival*. Cambridge: MIT Press.
- Calvert, R. 1987. "Reputation and Legislative Leadership." *Public Choice* 55: 81–119.
- Cox, G., and M. McCubbins. 1993. *Legislative Leviatan. Party Government in the House*. CA: University of California Press.
- Cox, G., and K. Poole. 2002. "On Measuring Partisanship in Roll Call Voting: The U.S. House of Representatives, 1877–1999." Unpublished Article. University of California at San Diego.
- Dal bo, E. 2007. "Bribing Voters." *American Journal of Political Science* 51: 789–803.
- Dekel, E., M. Jackson, and A. Wolinsky. 2006. "Vote Buying II: Legislatures and Lobbying." *Typescript*.
- Diermeier, D., and R. Myerson. 1999. "Bicameralism and its Consequences for the Internal Organization of Legislatures." *American Economic Review* 89: 1182–1196.
- Frankel, D., S. Morris, and A. Pauzner. 2001. "Equilibrium Selection in Global Games with Strategic Complementarities." *Journal of Economic Theory* 108: 1–44.
- Groseclose, T., and J. M. Snyder. 1996. "Buying Supermajorities." *American Political Science Review* 90: 303–315.
- Jones, C. O., Joseph G. Canon, and Howard W. Smith. 1968. "An Essay on the Limits of Leadership in the House of Representatives." *Journal of Politics* 30: 617–646.
- Kiewiet, R. D., and M. McCubbins. 1991. *The Logic of Delegation. Congressional Parties and the Appropriations Process*. IL: The University of Chicago Press.
- Krehbiel, K. 1993. "Where's the Party?" *British Journal of Political Science* 23: 235–266.
- Krehbiel, K. 1998. *Pivotal Politics. A Theory of U.S. Lawmaking*. Chicago, IL: University of Chicago Press.
- Lebo, M., A. McGlynn, and G. Koger. 2007. "Strategic Party Government: Party Influence in Congress, 1789–2000." *American Journal of Political Science* 51(3): 464–481.
- Lizzeri, A., and N. Persico. 2001. "The Provision of Public Goods under Alternative Electoral Incentives." *American Economic Review* 91: 225–239.
- McCarty, N., K. T. Poole, and H. Rosenthal. 2001. "The Hunt for Party Discipline in Congress." *American Political Science Review* 95: 673–687.
- McKenzie, R. 1964. *British Political Parties*. New York: Frederick A. Praeger Publisher.
- Michels, R. 1962. *Political Parties; A Sociological Study of the Oligarchical Tendencies of Modern Democracy*. Translated by Eden and Cedar Paul (first published 1911). NY: Collier Books.
- Morgan, J., and F. J. Vardy. 2006. "Corruption, Competition, and Contracts: A Model of Vote Buying." IMF Working Paper No. 06/11.
- Morgenstern, S., and B. Nacif (eds.). 2002. *Legislative Politics in Latin America*. Cambridge University Press.
- Morris, S., and H. S. Shin. 1998. "Unique Equilibrium in a Model of Self-Fulfilling Currency Attacks." *American Economic Review* 88: 587–597.
- Morris, S., and H. S. Shin. 2003. "Heterogeneity and Uniqueness in Interaction Games." Cowles Foundation Discussion Paper No. 1402, Cowles Foundation for Research in Economics, Yale University.
- Myerson, R. B. 1993. "Incentives to Cultivate Favored Minorities under Alternative Electoral Systems." *American Political Science Review* 87: 856–869.
- Myerson, R. B. 2005. *Leadership, Trust, and Constitutions*. Department of Economics, University of Chicago.
- Panbianco, A. 1988. *Political Parties: Organization and Power*. Cambridge: Cambridge University Press.
- Patty, J. W. 2008. "Equilibrium Party Government." *American Journal of Political Science* 52(3).
- Rohde, D. W. 1991. *Parties and Leaders in the Postreform House*. Chicago, IL: University of Chicago Press.
- Sinclair, B. 1983. *Majority Leadership in the House*. Baltimore: The Johns Hopkins University Press.
- Snyder, J. M. 1991. "On Buying Legislatures." *Economics and Politics* 3: 93–109.

- Snyder, J. M., and T. Groseclose. 2000. "Estimating Party Influence in Congressional Roll-Call Voting." *American Journal of Political Science* 44: 193–211.
- Tsebelis, G. 1995. "Decision Making in Political Systems: Veto Players in Presidentialism, Parliamentarism, Multicameralism, and Multipartyism." *British Journal of Political Science* 25: 289–325.
- Volden, C., and E. Bergman. 2006. "How Strong Should Our Party Be? Party Member Preferences Over Party Cohesion." *Legislative Studies Quarterly* 31(1): 71–104.