8

Strategic voting, party labels, and entry

Part II of this book investigated the nature and incidence of strategic voting in a variety of electoral systems, with particular emphasis on strategic desertion of weak candidates and lists. Duverger, and many after him, have argued that elite anticipation of strategic voting should lead to prudent withdrawals and hence a reduction in the number of competitors entering the field of battle. In particular, those elites who foresee that their own candidates or lists will bear the brunt of strategic desertion are likely to decide that mounting a (hopeless) campaign is not worth the cost, and seek instead to throw their support behind more viable candidates or lists (presumably for a price). To the extent that withdrawals of this sort do occur, the number of competitors will of course decrease.

In this chapter, I note that this argument about prudent withdrawals is theoretically limited in some of the same ways that the Duvergerian argument about strategic voting is limited. First, the argument presupposes that it will be clear *at the time at which entry decisions must be made* which candidate(s) or list(s) are doomed to be perceived as nonviable on the day of the election, hence shunned by instrumentally rational voters. So, just as voters must have consistent beliefs ("rational expectations") about who is trailing, elites must have consistent beliefs about who *will be* trailing and hence be the victim of strategic voting. Second, the prudent withdrawals argument also presupposes that elites are motivated primarily by the prospect of victory in the current election. Thus, just as voters must be short-term instrumentally rational, so must elites, for the argument to work.

In order to clarify these points, I first consider what happens when it is *not* clear, at the time at which entry decisions must be made, who will be perceived as nonviable. There are, of course, real-world situations in which such clarity is plausibly lacking: Think, for example, of elections

held in new democracies or in polities with unstructured party systems.[1] Here, I shall consider some formal models of entry that take things to the logical extreme: Every potential entrant is perceived (at the time entry must be decided) to have an equal chance of dodging the bullet of strategic desertion or, to put it the other way around, an equal chance of getting hit by it, should they in fact enter. To anticipate the result, such models show that the number of entrants is not limited by anticipations of strategic voting when everyone has an *ex ante* equal chance of suffering (or benefiting) from it. The only limits that are placed on the number of entrants in equilibrium have to do with the costs of entry and the benefits of office.

In Section 8.2, I consider another way in which Duverger's argument about prudent withdrawals might go awry, even if the issue of short-term viability is clear. I simply note various different kinds of possible long-term payoffs that might motivate elites, making them less likely than voters to coordinate.

In Section 8.3, I consider the case that is most favorable to the Duvergerian logic on entry: when all politicians are primarily concerned with doing well in the current election and a certain subset of potential entrants are clearly viable, with everyone else clearly nonviable. In this case, there is a sound reason not to enter if one is (consensually predicted to be) nonviable: One will probably face strategic desertion and lose badly, so why bother if the current election is all that one cares about? There will thus be an $M + 1$ rule at the level of candidate entry, if expectations are clear enough. But the real question is how it becomes so clear that large numbers of potential entrants have no chance of winning. Section 8.3 elaborates the role of party labels in coordinating expectations of viability and suggests that there may be an $M + 1$ rule governing the number of party labels as well.

Section 8.4 investigates the nature of entry in situations in which clear viability advantages exist for certain parties, due to their possession of valuable labels. Candidates value possession of these labels both because they convey a certain number of habitual voters into their camp and because they publicly certify the candidates' viability, thus insuring them against strategic desertion.

Given that candidates value labels with an established following in the electorate, at least two things follow. First, would-be career politicians will compete for the established labels that exist, as their chances of win-

[1]Sartori (1968:281, 293-4) defines a structured party system as one in which the established parties possess nationwide organizations and, more importantly for current purposes, command habitual allegiances in the electorate.

ning with such a label are typically far better than their chances as an independent or a new party's endorsee. Second, political groups, movements, and would-be parties will often find it more advantageous to join one of the viable parties, rather than setting up a new party of their own or joining a party that is not viable. In Section 8.4, I focus on the latter point, although of course the two are closely related.

8.1 NEUTRAL ENTRY MODELS

In this section I consider entry under conditions of extreme symmetry, with every potential entrant perceived, at the time of entry, as having as good a chance as any other to win. To capture such a situation, I turn to extant efforts in the literature based on a two-stage spatial model in which potential candidates first decide whether or not to enter the race and then, after some number have actually entered, compete by adopting positions somewhere along a left-right continuum (Palfrey 1984; Greenberg and Shepsle 1987; Ferejohn and Noll 1988; Feddersen, Sened, and Wright 1990; Weber 1990, 1992a, 1992b; Osborne 1993; Osborne and Silvinski 1995; Shvetsova 1995; Wada N.d.; see also Brams and Straffin 1982).[2] All such studies considered here employ the technical assumptions of neutrality and deterministic spatial voting, which together imply that the only characteristic of candidates that voters care about is their adopted left-right position, which voters know with certainty.[3] Assuming neutrality and deterministic voting means that voters are not "rationally ignorant" (Downs 1957) and accordingly need not utilize party labels as informational shortcuts (cf. Popkin 1991; Fiorina 1977) or make "standing decisions" in favor of candidates bearing a particular label (Key 1964b). Another way to put it is that candidates, and the parties that endorse them, have no spatial (i.e., ideological or policy) reputations to live down or build up. Assuming neutrality also means that voters have no cues, such as incumbency or

[2]Not all of the studies take exactly this form. For example, in Palfrey (1984) the entry stage is set up exogenously, with exactly two candidates entering; these two then take positions knowing that a third candidate will enter after they take their positions.

[3]A spatial model is *neutral* if, whenever candidates A and B switch spatial positions (all other candidates' positions held constant, if there are any other candidates), then they switch expected vote shares and probabilities of victory. The meaning of *deterministic voting* (as opposed to probabilistic voting; on which see Coughlin 1992) can most easily be explained in the case of two-candidate competition, say between A and B. If, whenever a voter prefers candidate A's policy position, then he or she votes for candidate A with certainty, then voting is deterministic. If A's superior spatial position makes it more likely that the voter will support A, but not certain, then voting is probabilistic.

past electoral history, that might tell them which candidates are viable and which are not. All the players are on a level playing field. This is in direct contrast to the more empirically-oriented decision-theoretic literature, in which the scare-off or entry-deterring effect of incumbency is often taken for granted.[4]

Four neutral entry models

In this section, I discuss four neutral entry models from the literature. The four models differ in various respects but all impose an upper bound on the number of entrants in equilibrium. The *existence* of an upper bound is reminiscent of Duverger's Law but, as will be seen, the *reasons* for the upper bound are quite different from what Duverger had in mind.

The first model I consider is that of Feddersen, Sened, and Wright (1990). In this model (of a simple plurality election), citizens vote strategically and a party can either enter a political contest at a specific spatial location, paying a cost of entry to do so, or decide not to enter at all. If a party sits the election out, it receives a payoff of zero. If, on the other hand, the party enters the fray, then its payoff equals its probability of winning (p), times the benefit of holding office (b), less the cost of entry (c). Thus, in equilibrium, no party will enter unless $pb \geq c$. In part because of the neutrality assumption, the equilibrium value of p turns out to equal $1/n$, where n is the number of entrants. That is, every entrant

[4]The decision-theoretic branch of the literature focuses on explaining the entry and exit decisions of individual candidates. Some, for example, investigate why incumbent legislators decide to retire or seek higher office, rather than seek reelection. Others investigate why nonincumbents seek legislative office to begin with. In both kinds of study, typical findings show that opportunity costs and chances of victory matter in a straightforward manner. As regards incumbents, for example, the literature shows that Republicans in the U.S. House of Representatives have sought higher office more frequently than their Democratic colleagues throughout the post-war era, because their party's perennial (albeit recently ended) minority status meant that they had little chance of attaining a committee or subcommittee chair (Gilmour and Rothstein 1993; Schansberg 1994); that U.S. Representatives from small states have sought statewide office more frequently than their large-state colleagues, because their House districts comprise a larger portion of the state electorate in which they must compete and they face less competition from other upwardly mobile House members (Rohde 1979; Brace 1984; Kiewiet and Zeng 1993:933); and that U.S. Representatives who kited more checks in the House banking scandal of 1991–2, plausibly reducing their (perceived) chances of reelection, were more likely to retire (Groseclose and Krehbiel 1994; Jacobson and Dimock 1994). As regards the scare-off effect of incumbency, Squire (1989:284) finds that only 4% of open seats in U.S. House elections were uncontested in the period 1978-88, whereas 14% of all seats were uncontested; Cox and Morgenstern (1993) find that the probability of a contested election in U.S. state legislative districts is from .02 to .14 greater if the seat is open than if it is defended by an incumbent (after controlling for relative party strengths in the district).

has an equal shot at winning in equilibrium. Thus, the condition $pb \geq c$ turns into $b/c \geq n$: *The number of entrants, n, is bounded above by the benefit-cost ratio, b/c.*

If one assumes that $b/c < 3$, then an upper bound similar to that claimed by Duverger's Law arises: In equilibrium, there will be at most two entrants. But the immediate reason for this equilibrium prediction is simply the assumption that $b/c < 3$. As Shepsle (1991:75; cf. Ferejohn and Noll 1988:15) puts it, "entry costs and office valuations drive the equilibrium number of entrants." But an emphasis on costs of entry and benefits of office has no obvious connection to Duverger's line of reasoning, which focused on elite anticipations of strategic voting. Feddersen, Sened, and Wright's approach puts one in mind of onerous signature requirements to get on the ballot in the United States, or the advantages given to "permanent" lemas in Uruguay (Gonzalez 1991), or the costs of television advertising in France (Duverger 1986:81). Duverger's approach centers around anticipations of strategic voting. Absent some reason to believe that the benefit-cost ratio *is* generally less than three, the Feddersen-Sened-Wright entry model does not lead to a result consistent with Duverger's Law, although it certainly does point to an important set of factors that might affect the equilibrium number of entrants in a given polity.

A second model of entry in simple plurality elections, due to Weber (1992b), differs from that of Feddersen, Sened, and Wright in a number of ways: parties maximize their share of the vote, not their expected utility; voters vote sincerely, not strategically; entry is sequential, not simultaneous. Yet the gist of the result on entry remains quite similar.

Weber posits that n established parties play simultaneously against one another, while anticipating the possible entry of a single potential entrant. The potential entrant will enter if and only if it can secure a share of the vote that exceeds a prespecified level, q. One interpretation of this is that the potential entrant derives utility from getting votes (possibly due to the probability that votes translate into seats, but possibly for other reasons as well), but also incurs a cost of entry. The level q is then simply the cost of entry expressed as a proportion of the total vote.

One result in Weber's model is that, if the cost of entry is sufficiently high (at least .25 expressed in share of vote terms), then two established parties will be able to deter entry by a third.[5] This prediction is the same as that made in Duverger's Law but, as with the Feddersen, Sened, and Wright result, depends purely on entry costs. Indeed, the absence of any reliance on elite anticipations of strategic voting is made even more obvious in this model by the stipulation that citizens vote sincerely.

[5]The mechanics of entry deterrence are similar to those in Palfrey (1984).

Strategic entry

A third entry model, due to Shvetsova (1995), also envisions a number of established parties playing simultaneously against one another, while anticipating the possible entry of a single additional candidate. Shvetsova's model, however, investigates entry under SNTV electoral rules in an M-seat district; the electoral structure is thus more general than that considered by Palfrey or Weber.

Shvetsova posits that all candidates maximize their probability of winning a seat, and that the potential entrant will enter if and only if it can secure a positive probability of winning. She also assumes that voters vote sincerely, not strategically.

One question Shvetsova asks (cf. Greenberg and Shepsle 1987) is whether an M-equilibrium exists: a situation in which M candidates have adopted spatial positions such that (1) these positions constitute a Nash equilibrium for the M prespecified entrants; and (2) the one potential entrant is deterred from entering. She also asks whether $M + 1$-equilibria, defined analogously, exist (cf. Weber 1992a). She finds that M-equilibria exist under stringent conditions (symmetric convex preference distributions and $M \leq 3$), that $M + 1$-equilibria exist for somewhat less stringent conditions (symmetric unimodal preference distributions), and that either an $M -$ or an $M + 1$-equilibrium must exist under yet less stringent conditions (unimodal preference distributions).

If one believes that preference distributions tend to be unimodal, then these results gibe to some extent with Duverger's Law and Reed's extension of it, in that the prediction is that there will be no more than $M + 1$ parties. The mechanism producing this result, however, has nothing to do with strategic voting, since voters in the model are assumed to behave sincerely. Instead, the result depends on the ability of the already-entered parties to collude, as it were, occupying all the attractive electoral niches and leaving no room for an $(M + 2)$nd entrant to eke out a positive probability of winning a seat.[6]

A final model that yields a result pertinent to Duverger's Law (Osborne and Silvinski 1995) compares entry under plurality rule and majority runoff. The model assumes that candidates value policy as well as the spoils of office and must pay a cost to enter the contest. In addition to finding that the number of candidates who enter (under either plurality or majority runoff) is sensitive to the cost of entry and benefit of winning, Osborne and Silvinski are able to show that the conditions under which

[6]Viewed more broadly, Shvetsova's result is essentially similar to those of Weber and Feddersen, Sened and Wright. Although the notion of a cost of entry is not explicitly introduced, her results can be interpreted as assuming that entry is costly, but that a positive probability of victory is sufficient to overcome this cost. (The model does not support arbitrarily small probabilities of victory in equilibrium, so in fact the assumed entry cost is not infinitesimal.)

a two-candidate equilibrium can be sustained under majority runoff are strictly more demanding than the analogous conditions for plurality rule. Thus, although the model does not support the conclusion that two-candidate equilibria are the only possible or likely equilibria under plurality rule (as Duverger's Law would suggest), it does show that two-candidate equilibria are more likely under plurality than under majority runoff. This result, however, has nothing to do with elite anticipations of strategic voting, since voters in the model are assumed to vote sincerely.

Discussion

The four models just reviewed all focus on the same basic mechanism as key in putting an upper bound on the number of entrants: Entry is costly and so new candidates will enter the fray only if their probabilities of victory (or vote shares) are large enough to justify the cost. This approach is not inconsistent with an emphasis on strategic voting: One can imagine building a model in which candidates estimate their probabilities of victory in part by anticipating who will bear the brunt of strategic voting. But in fact none of the models does this.

The only attempt in the literature to consider incorporating strategic voting as a factor influencing parties' perceptions of their chances of victory is due to Palfrey (1989), who merely notes (pp. 84-85) that a *neutral* entry model with strategic voting is "essentially indeterminate." If the model maintains neutrality, then before entry decisions are made, every party views itself as having as good a chance as any other in the position-taking subgame that will ensue post-entry. That is, no party has any reason to believe that *it* will be the likely victim of strategic voting, even though all believe that there will be victims. Thus, anticipations of strategic voting do nothing in a neutral entry model to change pre-entry (or even post-entry, pre-position taking) probabilities of victory, and hence do nothing to drive down the number of entrants.

How might an entry model incorporate elite anticipations of strategic voting? Taking the expected utility formulation of Feddersen, Sened, and Wright, the simplest way to incorporate anticipations of strategic voting would be via assumptions about the probability of victory. Parties would no longer calculate probabilities of victory, conditional on entry, by assuming voter neutrality and deterministic spatial voting. Instead, all parties would know that some of them had a nonspatial advantage, in that voters perceived them as more likely to be seriously in the running. The equilibrium result, in the voting subgame, would then be that disadvantaged candidates would likely bear the brunt of strategic voting and receive a zero vote share. Anticipating this, foreseeably disadvantaged candidates would not enter in the first place.

Thus, *if* it is clear at the time of entry who is viable and who is not, then entry by nonviable candidates should be deterred (to the extent that their entry is motivated by the chance of winning the current election). But how does it become clear who is nonviable? One answer is that it becomes clear from electoral history: Those who have won in the past become focal in any coordination games that ensue in subsequent elections (cf. Forsythe et al. 1993). The answer that I shall pursue later (Section 8.3) is that possessing the endorsement of a major party confers viability advantages.

8.2 WHEN POLITICIANS HAVE LONG-TERM PERSPECTIVES

Even if it is clear that a particular politician is likely to bear the brunt of strategic voting, were she to enter the race, she may still do so if she believes that entering will redound to her benefit in the future. Perhaps she believes that entering will cause a particular other candidate to lose the race, thereby demonstrating that her group's support is crucial and winning policy concessions. (This may have been the motivation of the Prohibitionist candidates in turn-of-the-century U.S. politics, for example.) Perhaps she believes that she will post a respectable showing, positioning herself for a more serious run at office next time. (Bernard Sanders, the lone Socialist in the U.S. House of Representatives, was able to establish credibility as a third-party candidate by first placing second, ahead of a weak Democratic candidate; cf. Endersby and Thomason 1994.) Perhaps she is a young British Conservative and understands that running well in a safe Labour district is a good way to get nominated in a better (more winnable) district next time.

If all politicians have long-term perspectives, then the coordination game between them is more complicated than Duverger implicitly assumed. Instead of being a one-stage coordination game, in which potential entrants jostle for position in the context of well-established expectations about viability, it is a multistage coordination game, in which an indefinite string of future elections is to be held, and expectations of viability may either be well-established (e.g., the Social Democratic Party (SDP) in the United Kingdom trying to supplant Labour) or not (e.g., a new democracy holding its first elections). In what follows, I shall focus on the case in which expectations are not well-formed.

In such games, it certainly makes sense to enter in the first round: By assumption, one has as good a chance as anyone else at winning seats and establishing focalness for later elections, or at least it is not clear that one does not. It also makes sense to "sound tough." If one can convince others that one is unalterably committed to entering in perpetuity, then others who seek to fill a similar niche in the field of candidates may be deterred from

entering. Just *saying* that one is committed, however, should not be very persuasive. The most obvious signal of commitment that has some credibility is to enter the first few rounds of competition, even if one's prospects are not good. The most natural equilibria of these kinds of multistage coordination games, then, are those in which there are a lot of entrants in the early rounds and a lot of huffing about long-term commitments.

From this perspective, the typical scenario in emerging democracies, whereby a great number of parties spring up in the first elections, and there is a relatively slow winnowing out process, makes sense. A large number spring up in the first election because it is not clear who will be viable and who not. As information is revealed about voter preferences, the more serious groups will continue to enter, even against poor short-term odds, in the hopes of convincing less committed competitors to drop out. The reward for such short-term sacrifice is establishing the clear expectation for future elections that one's group is the viable representative of a particular niche of electoral opinion.

8.3 ON THE VALUE OF PARTY ENDORSEMENTS AS COORDINATION DEVICES

If politicians are concerned primarily with the current election, and it is clear to all who the viable candidates are and who the nonviable candidates are, then coordination at the elite level will be easy and one expects no more than $M + 1$ candidates (or lists, in the case of PR) to enter. The $M + 1$ result at the level of strategic voting, in other words, induces an $M + 1$ result at the level of entry, if elites fully anticipate who will bear the brunt of strategic voting. The real question, then, is how "reputations for viability" are established or conferred. The answer in this section will be in terms of party endorsements.

Party endorsements can be valuable to voters in two main ways. The typical source of value noted in the literature is that endorsements can indicate, with variable precision, where the candidate receiving the endorsement stands on the issues (Key 1964b; Fiorina 1977). Another, less often noted, source of value lies in the endorsement's usefulness as a coordination device: If there were no nomination stage and no public endorsements, groups of like-minded voters might end up splitting their votes sub-optimally among a superabundance of similar candidates.

Party endorsements that are valuable to voters will for that reason also be valuable to candidates. To the extent that they convey information regarding the policy beliefs of candidates, and voters are rationally ignorant, they will carry along with them a certain mass of habitual followers. To the extent that the endorsement holds sway as a focal coordination device among some set of voters, it insulates endorsees from

strategic desertion, and, indeed, operates to make them a net beneficiary of strategic voting.

How can the value of a party's label be established or maintained? The label's value as a policy cue can be promoted by consistency and homogeneity of belief within the party. Then the party endorsement "means something" in terms of policy. The label's value as a coordination device is partly in the nature of a self-fulfilling prophecy. If all leftists believe that all leftists will cue on the Democratic party's endorsement, then it is rational for them to do so, rather than cueing on some other endorsement by some other group. Otherwise the Republicans, fewer in number but better coordinated, let us say, may steal the election.

Another way to put the last point is that a label must be focal in order to be valuable as a coordination device. It must have established a *monopoly* on endorsing within a given segment of opinion, having beaten out or coopted all the other would-be coordination devices.[7] If it has not beaten out the others, then voters are in the position of the subjects in Schelling's (1960) famous experiments, with too many bases on which to coordinate. Too many endorsements may thus be just as bad as none at all.

The preceding discussion has taken for granted that the end result of successful coordination is that the endorsee has a real shot at winning a seat. This in turn presumes a certain size. Successful coordination between a tiny minority of voters merely produces a voting bloc; it does not produce seats. Assuming that a payoff in seats is necessary to maintain a viable label, a maximum number of such labels is implied.

What this maximum is depends both on electoral structure and on how one defines viable. Consider first the main electoral systems considered in Part I, for which there are formally developed strategic voting models. There cannot generally be more than $M + 1$ candidates expected to be "in the running" for a seat in these systems, where M refers either to the district magnitude (in the case of plurality rule and proportional representation) or to the number of candidates that can qualify for the runoff (in the case of dual ballot systems). Defining a "viable" or "established" label to mean one whose top candidate has a shot at win-

[7]How *party* endorsements come to be preeminent, eclipsing endorsements by newspapers, prominent businesspersons, and so forth, is a fascinating question. My own hunch, based on a reading of the historical literature on the development of nomination procedures in the U.S., is that party nominations are constructed to be more persuasive about underlying strength in the electorate than are the endorsements that an interest group can issue.

ning every election, the maximum number of viable party labels competing in a given district is $M + 1$.[8]

Another way to think about this upper bound on the number of viable labels is in terms of the minimum viable size of a party. Certainly a party that can regularly attain or approximate the threshold of exclusion in a system will qualify as viable. As district magnitude increases, the threshold of exclusion, and hence the minimum viable size for a party, declines, so that the maximum number of viable labels increases.[9]

Moving beyond the purely electoral factors, being small will be less feasible if there are substantial economies of scale in purchasing television advertising, raising campaign funds, or securing government portfolios and other positions of power. There may thus be reason to expect that the number of viable *labels* will fall short, perhaps well short, of the number of viable *candidates*.

8.4 ENTRY IN THE PRESENCE OF ESTABLISHED LABELS

In the previous section, I argued that the maximum possible number of viable labels in a given district (operating under one of the electoral systems considered in Part I) would be $M + 1$. This upper bound comes essentially from a consideration of rational behavior on the part of voters under alternative electoral conditions.

Can one say anything more about the equilibrium number of labels, other than that it will not exceed $M + 1$? In this section, I imagine that there are a certain number of established labels, and consider the options of a "new" group that might wish to enter electoral politics. From an analysis of the group's decision in the face of these options, some further insight into the features of an equilibrium number of labels may be gained.

Assuming for present purposes that candidates running under established labels have an advantage (and not a disadvantage as recently in

[8]If a somewhat lower standard for viability were adopted, then of course the equilibrium upper bound on the number of viable labels would increase.

[9]Moving beyond the main electoral systems discussed in Part I, note that the minimum viable size of a party also declines with any other electoral manipulations – e.g., lowering legal thresholds, changing to a more proportional electoral formula, setting up upper tiers – that lower the threshold of exclusion. Being small is also made more feasible when electoral law facilitates either within-district vote trading (via *apparentement* or transferable-vote provisions) or joint nominating agreements (via joint listing or fusion provisions). In either case, parties below the minimum viable size can avoid paying any electoral penalty for their suboptimal size by allying with other parties. They can thus accrue the benefits of smallness-cum-consistency without sacrificing viability. Systems with intradistrict vote trading or joint nominating provisions thus may have more than $M + 1$ viable labels.

Italy), new political forces that seek to achieve power or influence in the short- or medium-term will be faced with a choice among four basic *electoral* options: (1) They can do nothing on the electoral front, perhaps discouraged by the high costs and low probabilities of success, and instead focus on nonelectoral strategies, e.g., lobbying the elected members of one or more of the established parties;[10] (2) they can seek to achieve influence over the endorsement process in one of the established parties, thereby securing safe list positions or safe constituencies for the candidates that they sponsor; (3) they can start a new party, motivated by a belief that they will be able to establish the new party's viability; or (4) even if they feel that there is no more room for another viable label in the system, they can become a "protest" or "blackmail" party, inflicting short-term damage on both themselves and one of the established labels, in the hopes of eventual concessions. In order to examine this four-way choice a bit more formally, it is useful to imagine a (unitary) group of some kind pondering its options in an M-seat district in which there are $m \leq M + 1$ established parties. To begin with, I shall consider only groups with very short time horizons, who would not consider the fourth option, nor the third if there was no *immediate* prospect of winning seats. Such groups thus choose among sitting the election out, seeking influence within one of the established parties, and running an independent or new-party slate of candidates.

I shall assume that each established party nominates at most M candidates, and that the nominees from party i can be ordered from 1 to n_i, where $n_i \leq M$.[11] The general interpretation of this ordered list of candidates (for which I shall use the term "slate") is that candidate 1 has a better (or at least no worse) endorsement position than candidate 2, who in turn is at least as favored as candidate 3, and so on. In closed-list systems of PR, the ordering corresponds simply to list position. In open- and flexible-list systems of PR (including STV), the idea that there is an ordered "slate" of names may have a natural analogue – in the Netherlands, for example, there is a party-determined order of names which substantially influences which candidates fill the list's seats – or it may not – in Finland, for example, I am not aware that the parties can differentially benefit their various candidates by allocating more and less favorable ballot positions; in this case, the formal notion of a "slate" has no real analogue and the ordering of candidates referred to above can be chosen at random. In Japan's system of SNTV, there is of course nothing that

[10]I shall assume that the group already pursues these nonelectoral stratagems at some optimal level, and is considering *adding* an electoral component to their overall strategy. Thus, the "do nothing" strategy can be taken to yield a utility increment of zero.

[11]Many of the world's polities limit the number of nominees in this fashion, but some do not.

would colloquially be called a "slate" but there may nonetheless be an advantage to having one's endorsement announced earlier rather than later. The Liberal Democratic Party (LDP), for example, typically announces the candidates that it will endorse in several rounds, and it may be advantageous to be in the first rather than second round. If so, then the formal notion of slate order would correspond to the chronological sequence in which nominations are announced. If, as is possible, order of announcement confers no systematic advantage, then the formal slate order can be chosen at random. In a single-member district, the issue of ordering does not arise.

Let e_{ij} be a dummy variable, equal to one if the group under consideration secures the jth position on party i's slate for one of its candidates, equal to zero otherwise. The vector $e_i = (e_{i1}, \ldots, e_{iM})$ then describes the overall success of the group in securing endorsement positions from party i, and the variable $e = (e_1, \ldots, e_m)$ describes the success of the group in securing endorsement positions from each of the m established parties. I shall assume that no group can secure endorsements from more than one established party. Thus, e_i is non-zero for at most one party i. If a group strikes out completely, whether through lack of trying or through being rebuffed, $e = 0$, where 0 is a vector of mM zeroes.

Let $s(e)$ be the expected number of seats that the group will win, given that it secures the endorsement positions described in e and runs a candidate in each secured position. Thus, for example, in a three-member Spanish district with two established parties, perhaps $e = (0,1,0;0,0,0)$ – the group gets the second spot on party 1's list – and $s(0,1,0;0,0,0) = .5$ – the second spot is competitive in this district. If $e = 0$, I shall interpret $s(0)$ as the expected number of seats the group will win if it runs an "optimal" number of candidates on an independent or new-party slate (the issue of optimal nomination arises, e.g., in limited vote systems).

The group cannot, of course, guarantee itself the endorsement positions that it wants. It is costly to enter party i's endorsement process and, even if the group can afford the costs of entry, it may not be successful in securing the spots it wants (or indeed any spots at all). To reflect these costs and uncertainties, let c_{ni} be the cost to the group of competing, at an optimal level of effort, in whatever process allocates nominations in party i, and let $p_i(e_i)$ equal the probability that the group will secure the endorsement positions e_i if it enters party i's nomination process and exerts its optimal effort.[12]

[12]As in the work of Black (1972) on progressive ambition, the expected utility calculation could be generalized to take explicit account of the effort level. As such refinements do not affect the basic points I am trying to make, I dispense with them here.

The decision problem facing the group can now be described in greater detail. The group must first decide whether or not to enter the endorsement process of one of the established parties and, if so, which one. If the group participates in an established party's process, then it must decide what to do next, in light of the endorsement positions that it secures. It may decide that the positions secured are not worth the costs of running in a general election or it may be denied any positions at all. In either of these cases, it would then face the problem of deciding whether to run as an independent or to sit the general election out. Alternatively, it may decide that the positions secured are good enough to be worth the costs of running in the general election, and proceed to do just that.

The expected utility of *not* entering *any* established party's endorsement process is expressed as follows:

$$\text{EU(Not Enter)} = \max\{s(0) - c_g(0), 0\} \quad (1)$$

where $c_g(0)$ represents the cost (denominated in seat-equivalents) of running an independent or new-party slate in a general election.[13] The logic behind this equation is simple: If the group under consideration does not compete for endorsement slots on the slate of any established party, then it will, come the general election, have a choice between launching an independent or new-party slate of its own, yielding an expected payoff of $s(0) - c_g(0)$, or sitting the election out, yielding a normalized payoff of zero. One might write the expected utility of running an independent or new-party slate as $s(0)b - c_g(0)$, where b is the value, in some unit of accounting, of a single seat. I have chosen the normalization $b = 1$, so that the cost function c_g is, as noted above, denominated in seat-equivalents.

The expected utility to the (risk-neutral) group of entering party i's endorsement process can be expressed formally as follows:

$$\text{EU(Enter } i) = \sum_{e_i \in E} p_i(e_i) \max\{s(e_i;0) - c_g(e_i;0), s(0) - c_g(0), 0\} - c_{ni} \quad (2)$$

where $E = \{0,1\}^M$ is the set of all possible sets of endorsement positions garnered by the group; the notation $(e_i;0)$ indicates that the group received no endorsement positions from parties other than i, while receiving the positions listed in e_i from i; and $c_g(e_i;0)$ represents the expected cost of running in a general election with the endorsement positions denoted by $(e_i;0)$.

[13] I assume that the focal group could run an independent or a new-party slate without bearing any costs analogous to the cost c_{ni} of entering party i's endorsement process; and I ignore the possibility of the group seeking to put its candidates on the slate of some other, already existing but nonviable, party.

The way to read equation (2) is as follows. For any given $e_i \in E$, the group under consideration has some probability $p_i(e_i)$ of securing the endorsement positions listed in e_i. Conditional on receiving those endorsement positions, the group can choose to sit the election out (yielding a normalized payoff of zero); to refuse the proffered nominations and run an independent or new-party slate (yielding a payoff of $s(0) - c_g(0)$); or to compete in the general election with the nominations garnered from party i (yielding a payoff of $s(e_i;0) - c_g(e_i;0)$). Regardless of which postnomination option the group chooses, it will have incurred the cost c_{ni} of participating in party i's nomination process, thus the last term on the right-hand side.

The group's decision between entering party i's nomination process and entering no process at all is governed by the sign of EU(Enter i) – EU(Not Enter): If this expression is positive, the group prefers entering i's process to entering none; if zero, the group is indifferent; if negative, the group prefers not entering any process to entering i's. Using equations (1) and (2), EU(Enter i) – EU(Not Enter) equals

$$\sum_{e_i \in C_i} p_i(e_i)[s(e_i;0) - c_g(e_i;0) - \max\{s(0) - c_g(0),0\}] - c_{ni} \qquad (3)$$

where $C_i = \{e_i \in E: s(e_i;0) - c_g(e_i;0) > \max\{s(0) - c_g(0),0\}\}$ is the set of all endorsement offers from party i that the group finds attractive, in the sense that it prefers running under party i's label in whatever positions it is given to its other options.

The entry decision as portrayed thus far focuses on current payoffs only. Future payoffs might be brought partially into view, albeit in a purely decision-theoretic way, simply by positing some discounted present value for the blackmail or protest option, and specifying continuation values for the other options as well. I will not develop this extension formally. The results stated below thus go beyond what is demonstrable in the confines of the formally developed model when they refer to the fourth option.

With that said, one can proceed to characterize the group's choice among the four options outlined above. The easiest case to consider is that of single-member plurality ($M = 1$) systems. In this case, $E = \{0,1\}$, $C_i = \{1\}$, and – assuming that $s(1;0) - c_g(1;0) - \max\{s(0) - c_g(0),0\} > 0$ – expression (3) reduces to $p_i(1)[s(1;0) - c_g(1;0) - \max\{s(0) - c_g(0),0\}] - c_{ni}$. From this expression we have:

Proposition 1: In single-member plurality systems, the probability that an office-seeking group will attempt to take over an established party's nomination process, rather than sit the election out, or enter the fray as a new party, increases with:

(a) The *permeability* of the major parties' endorsement processes (measured directly for party i by $p_i(1)$, the probability that the group can capture i's nomination for its own candidate, if it tries; and inversely by c_{ni}, the perceived costs of trying);
(b) The *advantage* of possessing one of the major parties' labels (measured for party i by $s(1;0) - c_g(1;0) - \max\{s(0) - c_g(0),0\}$, the expected value in seat-equivalents of the group's candidate campaigning *with* party i's endorsement, rather than *without* it).

The meaning of this proposition can be clarified by considering a few special cases that have already been noted in the literature.

Consider first the case where $s(1;0) \cong 1$, so that party 1 is dominant – indeed, virtually certain to win the general election – and the benefit of securing its label is maximal. In such a district, there is no (current seat-maximizing) reason to run under any other than the dominant label. Would-be career politicians will either enter the dominant party's endorsement process, or not at all. Groups that seek control of legislative power will similarly seek to take over the dominant party's endorsement process, or satisfy themselves with lobbying activities.[14]

These expectations of course correspond closely to conventional wisdom concerning the old "solid South" in the United States. There being no chance that a Republican might win, real political competition was diverted almost entirely into the Democratic primary (Key 1964a). As Epstein (1986:129) put the theoretical point: "Those who seek office [in one-party states] may perceive the primary of the dominant party as a more advantageous vehicle for success than entry, however easy, as candidates of a minority party. Protests, along with ambition, talent, and interest, are thus attracted to a single party." Grimm (1983:316) has noted a similar tendency in Germany for nomination contests to arise more frequently in districts that are "safe" for one or the other of the major parties (with the contest arising of course within the dominant party).

A separate claim often made about the South is that adoption of the direct primary helped to perpetuate one-partyism (cf. Epstein 1986:129-131). In terms of Proposition 1, this is a claim that a more permeable endorsement process – the direct primary as opposed to delegate caucuses or conventions – attracted more competition.[15]

Consider next the case in which $s(1;0) \cong s(0;1)$ and $s(0) - c_g(0) \leq 0$. This is a district in which two parties have a shot at winning the seat

[14]Blackmail parties of the kind described above will not be viable, because by assumption there is no chance that a third party could tip the election to the second party. There would thus be no pain inflicted and no reason to grant concessions.

[15]As an aside, it might be noted that when candidates *do* contest districts under a hopeless label, there are payoffs from doing so other than the prospect of winning the seat currently at stake or forcing policy concessions in the medium-term. In

but in which third parties and independents stand a sufficiently poor chance in the general election that it is not worth their effort to try (solely for the chance at the current seat). One result here is similar to that of the first case and depends on the substantial *advantage* of a major-party label: There is no (current seat-maximizing) reason to enter this district as anything other than one of the major-party candidates. Thus, new electoral forces (with short time horizons) will either try and force their way into one of the major's endorsement processes or sit the election out.

Another result in this case corresponds to Espstein's (1986:131-32) well-known argument that the widespread adoption of the direct primary in the United States, an eminently permeable endorsement process, contributes to the "distinctively American weakness of third parties." "The reasoning," Epstein notes, "is that third-party efforts are discouraged by the opportunity to capture the label of one or the other major party in a primary" (p. 131) and that "early abandonment of evidently failing third parties may be encouraged by the same ... opportunity that appears to make third parties less useful in the first place" (p. 132).

The logic underlying Proposition 1 is not limited to single-member plurality ($M = 1$) systems. A similar proposition holds for systems with $M > 1$ – whether top-M runoff systems in single-member districts, or PR and SNTV systems in multimember districts. The only complicating factor is that the permeability and benefit factors may not be nicely separable as in the $M = 1$ case. We may talk of $q = \sum_{e_i \in C_i} p_i(e_i)$ as the chance of getting a favorable set of endorsement spots, and of $B = \frac{1}{|C_i|} \sum_{e_i \in C_i} [s(e_i;0) - c_g(e_i;0) - \max\{s(0) - c_g(0),0\}]$ as the average advantage of a favorable set of endorsement spots, but we cannot simply multiply q and B to recover the expected benefit, as we could in the $M = 1$ case. Nonetheless, it is not too misleading to talk as if these terms were separable. If one does, then a more general version of Proposition 1 can be stated as follows:

Proposition 2: The probability that an office-seeking group will attempt to take over an established party's nomination process,

for example, contesting hopeless seats is an accepted means of demonstrating one's campaigning skills, and puts one in good stead for a better district next time. (The British example also points out a distinction between a label that is not viable in a particular district and one that is not viable anywhere. One reason that British third parties have trouble fielding good candidates in all districts is that they do not have enough safe and marginal constituencies to dangle as rewards for slogging into the more unpromising ones.)

rather than sit the election out, or enter the fray as a new party, increases with:
(a) The *permeability* of the established parties' endorsement processes (measured directly for party i by the probability q that the group can, if it tries, capture some good positions on i's slate for its own candidate(s); and inversely by c_{ni}, the perceived costs of trying);
(b) The *advantage* of possessing one of the major parties' labels (measured for party i by the average gain from having a favorable set of major-party endorsement spots, rather than running without them).

This proposition too can be clarified by considering some examples.

Consider first the cases of Uruguay and Colombia. Both elect their lower houses in multimember districts (ranging in magnitude from 2 to 29 in Colombia, from 2 to 47 in Uruguay).[16] Yet both have had long spells of democratic two-party politics (e.g., from 1974 to present in Colombia, from 1911 to 1971 in Uruguay). A reason for this similarity suggested by Proposition 2 is that in both countries the two major parties are very highly permeable.

Colombia is the simpler system to explain. All seats are allocated at the district level (no upper tiers or additional seats) using the largest-remainders method of PR with (usually) the Hare quota. There are, however, no legal restrictions on using the label of the two main parties, the Liberals and Conservatives. Anyone can run a list of candidates and call it a Liberal list, even if there are already Liberal lists in the field. The party label is thus a common good: exhaustible but nonexcludable. One result of this peculiar status of the major party labels in Colombia is that multiple lists bearing the same major-party label routinely appear in most districts. As votes for the various Liberal (or Conservative) lists do not pool for purposes of seat allocation, the system is similar to Japan's SNTV (cf. Cox and Shugart N.d.). Another result, related to the first, is that independent and third-party lists have been rare in Colombia (until the introduction in 1991 of a national list for the Senate, along with a new ballot format and party registration procedures, fueled an expansion). The rarity of such lists may have been partly due to the influence of the presidential race: Taking the Liberal (Conservative) label, rather than inventing some new one, associated a legislative list with a presidential candidate (hence, with the possibility of executive patronage and other favors).

The story in Uruguay is theoretically similar but more difficult to explain. Before the onset of military rule in 1971, a two-party system

[16]And both countries have long electoral and democratic traditions, albeit interrupted by civil war (in Colombia) and military rule (in Uruguay).

was maintained in Uruguay by, among other things, electoral laws that simultaneously ensured that both major parties would be highly *permeable* and that there would be a substantial *advantage* to running within an established party. The benefit of running within an established party (a "permanent" *lema*) was that all votes for lists running under the party label (under the *lema*) were pooled for purposes of seat allocations to the party, whereas no such advantage was offered to new groups. The advantage to running as a factional (or *sub-lema*) list within an established party (*lema*) was thus substantial.

Both major parties in Uruguay were also highly permeable. Both did maintain legal control of their labels but, precisely because all votes under the party label were pooled, party leaders had little incentive to deny the label to any group wishing to participate.

Colombia and Uruguay do not appear to be idiosyncratic. Shugart (1995) compared the effective number of legislative parties in seven countries (including Colombia but not Uruguay) that held their legislative and presidential elections concurrently. He found that those countries with more decentralized endorsement processes had significantly lower effective numbers of legislative parties. Shugart's interpretation of the data is that legislative candidates in concurrent elections are eager to align themselves with a presidential candidate, thus producing, in countries whose decentralized endorsement processes allow it, a reduction in the effective number of legislative parties.[17]

Although Propositions 1 and 2 speak of the permeability of a party's endorsement process, it is possible to interpret the theoretical parameters $\{p_i(e_i;0)\colon e_i \in C_i\}$ differently. Taken strictly, these are simply the probabilities that the group will secure various endorsements from group i. For an external group, these probabilities may well reflect the permeability or penetrability of the party. For a group that has already penetrated the party and is an active participant in it, however, these probabilities may reflect the balance of power within the party as much as or more than the permeability of the party.

Suppose one stipulates a relatively nonpermeable party and focuses on the *exit* decisions of intra-party factions. In this case, Proposition 2 suggests that a faction will exit whenever its likely endorsement placement(s) on the party slate yield fewer (expected) seats (net of costs) than would a run as an independent or new-party slate. The party as a whole, then, is more likely to hold together as (1) electoral life outside the party

[17]Shugart's evidence is not all positive. Looking at cases in which legislative and presidential elections were not held concurrently, he finds that countries with more decentralized endorsement processes had significantly *higher* effective numbers of legislative parties. It is not clear why this should be the case.

is less pleasant in prospect; and (2) there is less "underrepresentation" of factions in the endorsement process (whereby some factions take more than their proportional share of the endorsements, in view of their voting strengths, leaving others with less). These last comments may help explain why many of the world's stably factionalized parties either allocate endorsements proportionally to strength in the party conference (e.g., the Christian People's Party of Belgium, the Christian Democrats of Italy, the Union for French Democracy and Socialist Party of France) or "freeze" each faction's share (e.g., the Israel Labor Party in 1973).[18]

8.5 CONCLUSION

This chapter has considered several issues relating to political entry decisions and how they are conditioned by the anticipation of strategic voting. Duverger's original argument about entry in plurality elections was simply that would-be third-party candidates would anticipate their candidacies being ruined by strategic voting, and therefore not enter. The first task of this chapter has been to note some logical restrictions on the scope of this logic.

One restriction has to do with the clarity of expectations regarding viability. If, at the time at which entry decisions must be made, everyone is thought to be equally susceptible to strategic desertion, then no one in particular will be deterred from entering. It is only if it is clear, at the time of entry, who will probably suffer from strategic desertion that entry (by those unfortunates) will be deterred.

Another restriction has to do with politicians' goals. If they are concerned only with winning the current election, then the logic goes through. But if they are willing to suffer a string of losses in the hopes of eventual victory, or policy concessions, then it does not.

The force of the Duvergerian entry argument thus depends on whether these two conditions are generally met in a particular polity, or not. I do not have much to say about when politicians will have longer or shorter time horizons but I have suggested that generating clear expectations about viability is a matter mostly of electoral history and party labels. Once party labels have been established, and have the properties of (1) conveying a certain number of habitual votes into a candidate's total and (2) certifying the candidacy as "viable," then new candidacies that might compete for the voters that are aligned to the label are deterred. Such candidacies face a rather substantial coordination problem in competing with

[18]Japan's LDP has departed to a limited degree from this norm; see Cox and Rosenbluth (N.d.). If factions are risk-averse, then proportional allocation rules also provide some "insurance" benefits.

the nominee of the established label, with all of the advantages of focalness belonging to the nominee. If one views Duverger's Law as depending on two causal mechanisms – strategic voting, which reduces the effective number of candidates if more than two enter; and strategic entry, which reduces the number of candidates who actually enter to two – then the point can be restated as follows: The entry reduction part of the argument goes through if there is a structured party system in Sartori's sense of the term, but may not if there is not a structured party system.[19]

If indeed party labels are the primary devices of long-term coordination in most electoral systems, then the question arises of how many viable labels a given electoral system can sustain. Stating a unique optimal size for a party in a given electoral system is difficult. Being *small* is best for maintaining ideological consistency and making the party endorsement mean something in terms of policy. Being *big* is best for making a candidate or list focal and certifying that she, he, or it is viable; and for accruing any economies of scale in advertising, fund-raising, and government formation. At what size the overall profit from maintaining the label is maximized is hard to say.

One can say, however, something about what the minimum viable size of a party is. This depends more straightforwardly on thresholds of exclusion, which in turn are well understood functions of electoral structure (cf. Lijphart and Gibberd 1977). Dividing the minimum viable size into unity yields a maximum number of viable labels. This number is simply $M + 1$ for the three main electoral systems of Part II: plurality rule in M-seat districts, PR in M-seat districts, and top-M runoffs. In practice, the number of viable labels may well be smaller than this upper bound, if there are significant economies of scale in advertising, fund-raising, or government formation.[20]

[19]Sartori seems to take for granted that both parts of the district-level argument will go through in unstructured party systems. This may be true if there are other focal arbiters in the system: local oligarchs in prereform England, for example. But absent such alternative methods of certifying which candidates are viable (and which not), entry will not be deterred. So Duverger's Law would not operate fully even at the local level.

[20]This raises the question of what happens when the number of labels, m, falls short of $M + 1$, and new cleavages arise in the system. This is essentially asking whether an equilibrium in which $m < M + 1$ can be stable. Some possible responses to a new cleavage are compatible with maintaining the number of viable labels at m: if one of the major labels coopts the new issue, for example, or if activists seeking to push the issue onto the political agenda choose to infiltrate one of the major parties rather than start a new one (which decision would depend on the permeability of the various major parties' endorsement processes, *inter alia*). Another response, the creation of a protest or blackmail party designed to force issue cooptation or a favorable allocation of endorsements, suggests a short-term deviation from m, followed by a return. Finally, a third response, the creation of a new party, intended to be viable,

Strategic entry

If party labels confer viability and viability is otherwise hard to establish, then ambitious politicians who wish to win office may try to get a major party's endorsement rather than launch a new party or independent candidacy. The more valuable is running with a major-party label rather than without, and the more permeable is the major party's nomination process, the more likely are new groups or would-be candidates to "infiltrate" the major party, rather than start a new party, in accord with the old adage: If you can't beat them, join them.

increases the number of viable labels if successful (assuming that, because $m < M + 1$, there is actually room for another viable label and the creation of a new one will not simply displace an old one). Thus, when $m < M$, there is at least the logical possibility of the system accommodating some turbulence in electoral preferences without the creation of a new party. If, on the other hand, new cleavages arise in a system with $m = M + 1$, then the options are more limited. Cooptation, infiltration, and blackmail are still options but starting a new party with the idea that it will become viable is a much longer shot. If none of these strategies produce satisfaction, pressure to change the electoral system may mount, which can be considered a structural facilitation of new entry by would-be viable parties (e.g., New Zealand?). Of course, a desire to *consolidate* parties can also motivate electoral tinkering (as in Japan; cf. Christensen 1994).