

Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes to War

Author(s): George W. Downs and David M. Rocke

Source: *American Journal of Political Science*, Vol. 38, No. 2 (May, 1994), pp. 362-380

Published by: Midwest Political Science Association

Stable URL: <https://www.jstor.org/stable/2111408>

Accessed: 13-04-2025 22:15 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



JSTOR

Midwest Political Science Association is collaborating with JSTOR to digitize, preserve and extend access to *American Journal of Political Science*

Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes to War

George W. Downs, *Princeton University*

David M. Locke, *University of California, Davis*

The problem of ensuring that chief executives act in accordance with the wishes of their constituency is particularly acute in the area of foreign intervention where the head of state can be expected to possess substantial information advantages. This paper presents a formal analysis of strategies that can be used to deter overly passive and overly aggressive executives and a discussion of their side effects. The typically large amount of uncertainty means that the constituency must base its decision to retain an executive on the outcome of a conflict and not on its apparent *ex ante* advisability. This uncertainty imposes a cost on the constituency, who may remove an effective, "innocent" executive unnecessarily, and it also imposes a cost on the well-meaning executive, who may be removed from office after making the best possible decision in a difficult case. The mechanism necessary to deter executive adventurism also causes the paradoxical "gambling for resurrection" effect, in which an unsuccessful war that a well-informed principal would terminate is continued because cessation would, given the current state of the world, cause the agent to be removed from office.

There are two things that will always be difficult for a democratic nation: to start a war and to end it.

—Alexis de Tocqueville (1850, 649)

Introduction

It is axiomatic that the citizens of every state, democratic or not, have an interest in trying to ensure that their chief executive makes decisions that reflect their desires. This problem is particularly difficult in the areas of intervention and interstate conflict in which the chief executive is likely to have access to substantial amounts of information not available to his or her constituency and may possess preferences for or against participation in war that are different from those of the median constituency member.¹ In a democracy, the mechanisms that help deal with this

¹There seems to be more consensus in the literature about existence of differences in the preferences for conflict of constituencies and chief executives than in the direction of the difference. Those exposed to an autocratic system, like Thucydides and Kant, appear to believe that leaders frequently have more aggressive preferences than their subjects. Those, like Lippman (1955) and Kennan (1951), whose experience lies more in democracies have argued the opposite. Our argument requires nothing more than the real possibility that a difference exists.

principal-agent problem range from a free press and legislative declaration of war to electoral defeat and impeachment. In an autocracy, there are far fewer of these mechanisms, and at the extreme there may be nothing more than the costly option of armed rebellion.

We present a formal analysis of how a constituency should employ the sanction of removing an executive from office—through electoral or forcible means—to cope with the information advantages of a chief executive, and we describe the negative side effects that are associated with this strategy. This particular mechanism is important for a number of reasons. First, while the costs of imposing it vary greatly across societies, it is usually thought to be the most widely available and effective sanction for controlling executive behavior. As expressed through the electoral option in democratic societies, it has been found to play an important role in shaping the foreign policy behavior of governments (Gaubatz 1991; Morrow 1991). Second, the removal of a chief executive by whatever means is likely to provide subsequent executives with the clearest signal about the limits of citizen tolerance and the character of citizen preferences. The expectation that this signal creates is a critical part of any agency solution. Third, removal is the option of last resort that can be employed when other mechanisms fail. Thus, it can be used in circumstances in which the chief executive has concealed information or has acted outside the bounds of due process.

One important implication of the model is that high information uncertainty forces a citizenry to gauge by battlefield success or by the apparent consequences of inaction the extent to which an executive is acting in a manner that is consistent with its preferences. In particular, it calls for the removal of an executive who initiates—or actively perpetuates—a losing campaign. This sanction is invoked not as punishment for failure or out of confusion about the nature of sunk costs but to deter subsequent executives from exploiting their information advantage to wage aggression when a fully informed citizenry would not. To learn whether real-world behavior corresponds to prescriptions of the model, we examine the events that follow unsuccessful aggression by a set of primarily non-democratic states. Although the high cost associated with the removal of the executive in nondemocratic states will work against the hypothesis, these are states whose citizens are likely to have the poorest quality information and the fewest alternative mechanisms of agency control.

A second implication involves the cost and negative side effects associated with this inevitably second-best solution. One is that information asymmetries create a situation in which it is extremely difficult to distinguish incidents in which a good decision has led to a bad outcome from incidents in which a bad decision has led to a bad result. As a consequence,

it is inevitable that a system will be established where some percentage of chief executives will be removed from office—indeed, have to be removed to prevent excessive aggressiveness—despite the fact that they have acted in a fashion that was entirely consistent with the public interest.² Thus, for a “good” public-regarding executive, the privacy of information is more of a curse than a blessing. Another problem is that removal from office is a crude, ungraduated instrument that brings with it the problem of limited liability. Once a chief executive believes that removal is likely, he or she has nothing more to lose by “gambling for resurrection” by escalating or extending a conflict. Finally, the distinction between an executive who is more aggressive in initiating conflict and one who is gambling for resurrection is often blurred to the point where a state suffering the effects of the first may attempt to reduce the chance of that problem recurring. This will indeed result in fewer conflict decisions, but the incentive to gamble for resurrection will increase.

Both of these implications distinguish what is a hidden information problem from the typical retrospective voting model.³ In the latter models, greater or lesser voter concern for the economy might affect the incentive for elected officials to attempt to signal their competence by artificially manipulating economic performance, but the failure of the voters to employ a decision rule that distinguishes competent from incompetent executives does not increase the chances that the next executive will act incompetently. In the case of the conflict agency model, however, a failure to punish one executive will increase the likelihood that subsequent executives will disregard constituency preferences. In the domestic economy agency model, gambling for resurrection behavior is not modeled at all, although it arguably should be. One can easily imagine circumstances where an executive, competent or incompetent, who has the misfortune to be caught in a poor economy will be tempted to gamble for resurrection by implementing high-payoff, low-probability policies.

We begin by constructing a simple model that represents, in sche-

²In spite of the fact that uncertainty and private information should be less prevalent in domestic affairs than in the international arena, results of Alesina, Londregan, and Rosenthal (1993) suggest that this phenomenon occurs in the former area as well. They found that voters apparently are influenced by the actual outcome of economic growth and not by the difference between a reasonable prediction of the outcome and what occurred. One possible explanation of this is that voters use this sanction to ensure that manipulation of expectations cannot be used by the executive to make herself or himself look good.

³Both the formal retrospective voting literature and the domestic political economy literature are too rich to be summarized here. See Persson and Tabellini (1990), Rogoff and Sibert (1988), Hibbs (1987), and Norpoth, Lewis-Beck, and Lafay (1991).

matic form, a game between a constituency and the chief executive, that commences when a conflict opportunity arises. The executive possesses information about the probable benefits, costs, and outcome that the constituency is not privy to and may well remain ignorant of even after the resolution of the crisis. The task of the constituency is to create a decision rule for the use of its removal sanction that deters, to the extent possible, its leader from misrepresenting the situation in such a way that conflict is initiated when an accurately informed constituency would want it to be avoided, or vice versa. Moreover, this decision rule should be constructed in such a way that it does not distort the actions of an executive who would already be inclined to do as the constituency wanted.

The model is intentionally one dimensional in that it assumes that the constituency is preoccupied with ensuring that the executive will act as the constituency itself would with respect to initiating a conflict. This is done to understand as completely as possible the effect that uncertainty has on the exercise of the removal sanction. Obviously, to the extent that other issues are more important and success in other areas can compensate for failure in this area, the model will be incomplete.

The Theory

Assume that there is a situation that potentially calls for the initiation of conflict (e.g., an international crisis). The executive possesses information θ that predicts the outcome (the success or failure and the associated costs and benefits) more precisely than is generally known. This provides the executive with the opportunity to conceal or distort for private purposes (e.g., drawing attention away from domestic problems, imperial ambitions, extreme risk averseness) information that would affect the public estimate of the advisability of conflict. Conversely, the existence of θ provides the constituency with a motive to establish some incentive scheme to prevent this.

The sequence of events in the model begins with the arrival of the crisis. The public has a best estimate in the form of prior means for the outcome X (expressed in units of the public's utility) that results from war or peace. The executive then observes the private information θ that will, in general, shift the best guess (mean) of the future value of X and also lowers the variance. Then the executive decides whether to initiate conflict, and the outcome is observed. Based on the outcome and a comparison with prior expectations, the public decides whether to remove the executive. The goal of the constituency side of the game is to devise rules for applying the sanction based on X that give an incentive for the executive to behave as the public would if it too had observed θ .

Formally, we assume that dependence of the distribution of X on the parameters θ (the private information) and I (an indicator of whether conflict has been initiated) is in the form of location and scale shifts of a single underlying distribution F ; that is,

$$F(x|\theta, I) = F\left(\frac{x - \mu_{\theta I}}{\sigma_I}\right), \quad (1)$$

where $\mu_{\theta 1}$ and $\mu_{\theta 0}$ are the means of X under war and peace, respectively, and where we suppose for simplicity that the variance is constant at σ_1^2 and σ_0^2 , respectively. We shall scale θ so that a positive θ indicates that conflict will have stochastically better outcomes than inaction and negative θ that it will have worse outcomes. Specifically, we assume that

$$\mu_{01} = 0, \quad (2)$$

$$\mu_{00} = 0, \quad (3)$$

$$\frac{\partial \mu_{\theta 1}}{\partial \theta} > 0 \ (\forall \theta), \quad (4)$$

$$\frac{\partial \mu_{\theta 0}}{\partial \theta} < 0 \ (\forall \theta), \quad (5)$$

$$\sigma_1 > \sigma_0. \quad (6)$$

The first two conditions are merely a scaling effect, but the second two express the fact that the expected payoff to initiating conflict (inaction) is an increasing (decreasing) function of θ . The final condition says that outcomes of war are less predictable, in the sense of having a higher variance, than outcomes of inaction.

If the constituency knew θ , then they could use the replacement sanction to force the executive to adhere to their preferences by simply declaring beforehand that he or she must engage in conflict whenever a positive θ is observed and refrain from doing so otherwise. To the extent that the executive valued continuance in office more than the utility of engaging in the behavior that the constituency did not want, this strategy would be optimal. Unfortunately, under the more typical situation where the executive has θ as private information and there is uncertainty about the extent to which his or her preferences mirror those of the median constituency member, it is necessary to devise a good second-best strategy.

Let α be a parameter that describes the tendency of the executive toward overaggressiveness, which we shall denote for simplicity by the term *adventurism*, or overcautiousness, which we denote by *timidity*.

Suppose that $\alpha = 0$ corresponds to the executive having the same preference structure as the constituency with $\alpha > 0$ showing a preference for adventurism and $\alpha < 0$ showing an overly cautious attitude (compared to the constituency).⁴ Thus, the executive maximizes

$$E(u(X; \alpha) | \theta = \theta_0, I) \quad (7)$$

over choices of $I = 0$ or 1 .

The primary purpose of introducing a removal sanction is for the constituency to force the executive's behavior to be nearer their own preferences than would be the case without the sanction. In this model, the only item of information that the constituency can use to decide whether the executive should be removed from office is the outcome X of the crisis that led to war or peace. Clearly, the only reasonable form this can take is that, if war has occurred and $X < E_1$, the executive is removed; if war has not occurred, and $X < E_0$, then the executive is removed. The problem facing the public is to choose the points E_0 and E_1 at which removal results.

The public might reasonably have three primary goals that would help determine good values of E_0 and E_1 . The first is that if the executive is unduly adventuresome ($\alpha > 0$), then a disincentive to initiate conflict would be applied. This should cause some decisions to initiate conflict—incorrectly made from the public's point of view—to be switched to correct ones. Second, if the executive is too timid ($\alpha < 0$), then a disincentive to hold back should be applied. This should cause some incorrect decisions to hold back to be switched to correct ones. Finally, the incentives should preferably not ever cause what would have been a correct decision to be switched to an incorrect one.

Technically, it is easier to examine first what would be necessary to ensure the third condition. If no correct decision should be turned into an incorrect one by the application of sanctions, it is necessary that a greater disincentive be applied to conflict initiation when $\theta < 0$ and a greater disincentive to hold back when $\theta > 0$. By continuity, this means that there should be no incentive either way when $\theta = 0$ because then the public would be indifferent between war and peace (if θ were known). In particular, this means that an executive with the same preferences as

⁴The parameter α may index attitude toward risk, with negative α indicating a preference for low-variance outcomes. Alternatively, α may index an intrinsic value placed on initiating conflict by the executive. For what follows, this distinction is unimportant—it is only necessary that α index the degree to which the executive would initiate conflict when the public would prefer not to.

the public ($\alpha = 0$) should have no decision changed by the potential imposition of sanctions. A sufficient condition for this to occur is that, when $\alpha = 0$ and $\theta = 0$, the executive is still indifferent between war and peace even considering the incentive effects of sanctions.

Since there is only one sanction of unvarying size (which we denote by $-Q$), the only way to ensure incentive neutrality in this case is to require that the probability of imposing the sanction after initiating conflict is the same as the probability of imposing the sanction under inaction, when $\theta = 0$. Technically, if F is the distribution function of X , then this condition is

$$F(E_0|\theta = 0, I = 0) = F(E_1|\theta = 0, I = 1). \quad (8)$$

This implies that there is only one parameter p that the constituency may choose and that

$$E_0 = F^{-1}(p|\theta = 0, I = 0), \quad (9)$$

$$E_1 = F^{-1}(p|\theta = 0, I = 1). \quad (10)$$

If we let $z_p = F^{-1}(p)$, then

$$E_0 = \mu_{00} + \sigma_0 z_p, \quad (11)$$

$$E_1 = \mu_{01} + \sigma_1 z_p. \quad (12)$$

This illustrates a specific difficulty that the constituency has in this problem. In a typical principal-agent model,⁵ the principal can choose an action from a continuum (amount of compensation), whereas here the constituency has only a binary sanction. They cannot choose the intensity of the punishment—only whether it occurs or not. Ex ante, this means that the parameter the constituency can choose is the probability that the sanction will be imposed.

To appreciate how the use of a sanction as described above provides a deterrent for either excessive adventurism or excessive timidity, we can calculate what the critical level of θ is for initiating conflict both with and without the sanction. Without the sanction, the executive will initiate conflict whenever

$$E(u(X; \alpha)|\theta, I = 1) - E(u(X; \alpha)|\theta, I = 0) > 0, \quad (13)$$

whereas with the sanction the condition is

$$\begin{aligned} & (E(u(X; \alpha)|\theta, I = 1) - QF(E_1|\theta, I = 1)) \\ & - (E(u(X; \alpha)|\theta, I = 0) - QF(E_0|\theta, I = 0)) > 0. \end{aligned} \quad (14)$$

⁵See Fudenberg and Tirole (1991), Levinthal (1988), or Tirole (1990).

The left-hand side of equation (14) is less than the left-hand side of equation (13) by

$$QF(E_1|\theta, I = 1) - QF(E_0|\theta, I = 0). \quad (15)$$

When $\theta < 0$, so that the constituency would prefer inaction, this quantity is positive (by equations 2–5), so that a disincentive to initiating conflict is provided. Similarly, when $\theta > 0$, so that the constituency would prefer conflict, a disincentive is provided to inaction.

The above analysis is valid for any value of p , the probability of applying the sanction at the indifference point $\theta = 0$. There still remains the issue of optimal choice of p from the constituency point of view.⁶ To see how this should be selected, we shall begin by examining the sequence of events in the game, which is as follows:

1. The constituency chooses p and therefore z_p .
2. The information θ is observed by the executive. Suppose that the ex ante distribution of θ is a known function $G(\cdot)$.
3. The executive decides whether to initiate conflict. This depends on the type α of the executive, information θ , and the sanction variable p .
4. The constituency observes a utility value X , which depends on θ and the decision of the executive. It may also depend on a prior distribution $A(\cdot)$ that the constituency has for the type of the executive.

The complexity and conditional nature of this sequence of events makes an exact, closed-form analysis difficult. The approach we take in dealing with it is to identify a plausible value of θ at which the constituency would wish to provide the maximum disincentive to intervention and then to use the optimal p for this value of θ . In this form, the optimal p does not depend on α , since the election sanction is assumed to be a separable contribution to the utility of the executive. This does not amount to an advance assumption about the value of θ . Rather, it provides a value of θ at which to standardize. This value of θ would correspond to a situation that might plausibly occur and in which war would be particularly disastrous. An alternative, which we explore in the appendix, is to make more complex demands on the model, which results in implicit rather than explicit solutions. There, we shall analyze the case

⁶That is, optimal given the information constraints. Obviously, any policy formulated without benefit of all the information is second best; under the circumstances, this is the best that can be done.

where the constituency has an (accurate) prior distribution on the values of α and θ .

If $\theta > 0$, then the issue is to deter inaction, and if $\theta < 0$, then the issue is to deter the initiation of conflict by the executive. If we choose the optimal policy for the utility at expectation of θ , rather than the exact value, which is the optimal p for the expected utility at θ , then for a given value of $\theta < 0$, the optimal choice is the one that maximizes

$$\begin{aligned}
 D(p, \theta) &= Q(F(E_1|\theta, I = 1) - F(E_0|\theta, I = 0)) \\
 &= Q\left(F\left(\frac{E_1 - \mu_{\theta 1}}{\sigma_1}\right) - F\left(\frac{E_0 - \mu_{\theta 0}}{\sigma_0}\right)\right) \\
 &= Q\left(F\left(\frac{\mu_{01} + \sigma_1 z_p - \mu_{\theta 1}}{\sigma_1}\right) - F\left(\frac{\mu_{00} + \sigma_0 z_p - \mu_{\theta 0}}{\sigma_0}\right)\right) \\
 &= Q\left(F\left(z_p + \frac{\mu_{01} - \mu_{\theta 1}}{\sigma_1}\right) - F\left(z_p + \frac{\mu_{00} - \mu_{\theta 0}}{\sigma_0}\right)\right).
 \end{aligned} \tag{16}$$

The first-order condition for the maximum value of D with respect to the choice of p is given by

$$\begin{aligned}
 0 &= f\left(z_p + \frac{\mu_{01} - \mu_{\theta 1}}{\sigma_1}\right) z'_p - f\left(z_p + \frac{\mu_{00} - \mu_{\theta 0}}{\sigma_0}\right) z'_p \\
 0 &= f\left(z_p + \frac{\mu_{01} - \mu_{\theta 1}}{\sigma_1}\right) - f\left(z_p + \frac{\mu_{00} - \mu_{\theta 0}}{\sigma_0}\right),
 \end{aligned} \tag{17}$$

and the second-order condition is

$$\begin{aligned}
 0 &< f'\left(z_p + \frac{\mu_{01} - \mu_{\theta 1}}{\sigma_1}\right) (z'_p)^2 - f'\left(z_p + \frac{\mu_{00} - \mu_{\theta 0}}{\sigma_0}\right) (z'_p)^2 \\
 &\quad + f\left(z_p + \frac{\mu_{01} - \mu_{\theta 1}}{\sigma_1}\right) z''_p - f\left(z_p + \frac{\mu_{00} - \mu_{\theta 0}}{\sigma_0}\right) z''_p.
 \end{aligned} \tag{18}$$

Clearly, some conditions are required on F for the existence of a solution to be guaranteed. The theorem below shows that sufficient conditions are that the density f is differentiable and strongly unimodal.

THEOREM. Let the standardized density f of X be differentiable and strongly unimodal. Then, for each fixed θ , there exists a unique solution to the constituency's maximization problem defined by equation (16).

PROOF. Strongly unimodal means that there exists a unique point M such that $f(x)$ is strictly increasing for $x < M$ and strictly decreasing for $x > M$. Without loss of generality, we shall take $\alpha > 0$ and $\theta < 0$; the other cases follow directly. In this case, $-a = (\mu_{01} - \mu_{\theta 1})/\sigma_1 < 0$ and $b = (\mu_{00} - \mu_{\theta 0})/\sigma_1 > 0$. Choose p so that $z_p + b = M$; then the first-order criterion is clearly negative. Choose p so that $z_p - a = M$; then the first-order criterion is clearly positive. By continuity, there exists a zero. Strong unimodality implies that there cannot be a second solution. If both $z_p - a$ and $z_p + b$ are on the same side of the mode, then the density values cannot be equal, since the density is strictly decreasing (increasing). If they are on opposite sides of the mode, as is the case with the solution just derived, then the difference of the densities is clearly strictly increasing and so cannot have another zero. The second-order condition (18), evaluated at the solution, becomes

$$f'\left(z_p + \frac{\mu_{01} - \mu_{\theta 1}}{\sigma_1}\right)(z_p')^2 - f'\left(z_p + \frac{\mu_{00} - \mu_{\theta 0}}{\sigma_0}\right)(z_p')^2, \quad (19)$$

which is clearly positive (because $z_p - a$ and $z_p + b$ lie on opposite sides of the mode), showing that the critical point is a maximum. QED

What this theorem shows is that given a particular situation that is either likely or greatly feared, where the executive needs to be deterred from a particular act of adventurism (timidity), there is an optimal strategy to do so. It is not possible optimally to deter undesired actions in all situations, so one particular type must be chosen as being likely to occur, or important to deter, and then the optimal policy for this case is used in general.

As a specific mathematical illustration, suppose that the ex ante distribution of θ is normal with mean $-2\sigma_0$, so that the outcome would almost always favor inaction rather than war. Assume that $\mu_{\theta 1} = \theta$ and that $\mu_{\theta 0} = -\theta$ (this is a simple way to have θ indicate whether conflict initiation is desirable) and suppose that F is standard normal. Then the condition for optimality becomes

$$\begin{aligned} \left(z_p - \frac{\theta}{\sigma_1}\right)^2 &= \left(z_p + \frac{\theta}{\sigma_0}\right)^2 \\ -\frac{2z_p\theta}{\sigma_1} + \frac{\theta^2}{\sigma_1^2} &= \frac{2z_p\theta}{\sigma_0} + \frac{\theta^2}{\sigma_0^2} \\ z_p &= \frac{\theta^2(\sigma_1^{-2} - \sigma_0^{-2})}{2\theta(\sigma_1^{-1} + \sigma_0^{-1})} \\ &= 0.5\theta(\sigma_1^{-1} - \sigma_0^{-1}). \end{aligned} \quad (20)$$

Standardizing at $\theta = -2\sigma_0$, one would use

$$\begin{aligned} z_p &= -\sigma_0(\sigma_1^{-1} - \sigma_0^{-1}) \\ &= 1 - \frac{\sigma_0}{\sigma_1}. \end{aligned} \tag{21}$$

If war had twice the standard deviation of inaction, then one would choose $z_p = .5$, so that $p = .5$, $E_0 = E_1 = 0$. Note that the assumption that war has a higher standard deviation is necessary here; otherwise all choices of p are equally preferable.

The closed-form solution (20) allows us easily to comment on the dependence of the solution on the parameters. Note that z_p is a linear function of the situation at which the punishment is to be standardized. As θ becomes more and more negative, so that war would be more and more undesirable, the value of z_p rises, and the optimal percentage of the time that the sanction is applied also rises. The other critical variable is the ratio σ_1/σ_0 between the standard deviation of the result under war and the standard deviation under peace. The closer this is to one, the less likely is the sanction to be applied.

Empirical Corroboration

The bottom line from a qualitative standpoint is that, under information asymmetry, a constituency that places a high value on having its wishes respected with regard to conflict initiation will key its use of the removal sanction to conflict outcomes. *Ceteris paribus*, the greater the information asymmetry, and the worse the conflict outcome, the more we expect the removal rate of executives who are unsuccessful at war to exceed the removal rate of executives who do not go to war.

Evidence that constituencies act in a fashion consistent with the model can be found in Bueno de Mesquita, Siverson, and Woller (1993). Using data that consisted of both war-participating and nonwar-participating states between the years 1816 and 1975, they find (1) that leaders who initiate and lose a costly war stand an 80% chance of being replaced and (2) that the probability of replacement increases with the severity of the defeat (measured in battle deaths).⁷ Most of the regimes in question were autocracies, where we would expect the information asymmetries to be the greatest.⁸

⁷Cases where the regime change was forced by an enemy were dropped from the analysis.

⁸Most autocracies as well as democracies are vulnerable to the agency problem described here, since executives of autocratic states are, at least at some point, subject to forcible removal by an elite that functions in the role of the principal. For a related application of agency theory to the diversionary use of force, see Richards et al. (1993).

Obviously, these findings are more suggestive than conclusive. A proponent of competency-based retrospective voting, for example, could argue that the findings indicate nothing more than an attempt to remove incompetent, rather than adventuresome, executives. While we think that competency assessment may well be playing a role, Bueno de Mesquita, Siverson, and Woller (1993) also found that leaders who initiated a losing war were removed from office at a higher rate than those who lost a war that they did not initiate. The argument that initiators are more adventuresome than noninitiators seems somewhat stronger than the argument that those who initiate losing wars are more incompetent than those who fail to prepare for them. Furthermore, a competency-based removal would not be necessary for executives who are not subject to reelection, since the cost of the executive's mis-estimation is sunk and cannot be recovered by punishment. Thus, our agency model seems to provide the best explanation. A further fit between our model and the results of Bueno de Mesquita, Siverson, and Woller (1993) is that our model predicts that the worse the outcome, the greater the likelihood of removal, since removal is a function of how bad the outcome was and not just that it was bad. This phenomenon was observed by Bueno de Mesquita and his colleagues.

The Executive's Dilemma

Second-best solutions to agency problems are not infrequently costly from the agent's perspective as well as from that of the principal. The irony in this case is that a cost is borne by executives whose preferences are precisely what the constituency would like them to be, but who have the misfortune to see their responsible decisions in difficult situations result in poor outcomes.

We have seen that, because the constituency cannot observe directly whether the executive has made the right decision, it must establish a set of punishments for poor outcomes or remove the deterrent to future executives who might be tempted to act contrary to the constituency's interests. Unfortunately, the relationship between the quality of a decision and the character of the outcome is notoriously imperfect. Inevitably, a certain percentage of decisions that perfectly represent constituency desires *ex ante* will lead to bad outcomes, and the executive responsible will be punished for no other reason than that he or she was unlucky and the constituency does not possess the information to distinguish bad luck from manipulation.

In this model, the parameter p is the chance that an executive whose preferences perfectly mirror those of the constituency will be turned out of office if faced with a cusp decision where $\theta = 0$. The constituency must do this to deter executives with different objectives. This has the

interesting implication that the worst thing that can happen to an executive is to be faced with a close decision. If the information θ clearly indicates a preferred alternative, then the executive can act so as to reduce the chance of being expelled from office far below p by acting as the public wishes. If, however, the information θ is completely neutral, then either way the executive acts will result in the (perhaps large) chance p of removal from office. Note that this is different from saying that close decisions are difficult to make. The assertion is that they are dangerous to the executives' political health, whichever way the decision goes.

Gambling for Resurrection

The models we have developed so far in this paper are schematic in a number of ways. One of the most important is that war actually is not a once-for-all decision—rather, it unfolds across time, with many opportunities to continue or reduce the level of conflict. Further, a decision not to initiate conflict at one moment does not imply that a decision to act cannot be taken later. Moreover, the sanction of removal from office cannot be applied instantaneously. In the United States, the voters' sanction can be applied only at fixed election times. Even in a parliamentary democracy, there are inevitable time lags between a given event, public disapproval, the decision to call an election, and the election itself. Practical obstacles are even greater in autocracies. The period of time covered by the model should be thought of as the interval between the initiation of conflict and the first opportunity to remove the executive.

Consider what happens now if we divide the period of executive decision into two stages. We have the following sequence of events:

1. The executive obtains a piece of private information θ_1 that gives the executive improved estimates $\mu_{\theta_1 I_1 1}$ ($I_1 = 0, 1$) of the average result of conflict or inaction in the first stage.
2. The first stage outcome X_1 is observed by the executive and may or may not be observed by the constituency.
3. The executive obtains a second piece of private information θ_2 that gives the executive improved estimates $\mu_{\theta_2 I_2 2}$ ($I_2 = 0, 1$) of the average added result of conflict or inaction in the second stage.
4. The second stage outcome X_2 is observed by the executive and possibly the constituency. The total outcome (in units of the constituency's utility) is $X = X_1 + X_2$, which is observed by all.
5. The constituency decides whether the executive (or the party of the executive) should be kept in office based on the two decisions and on the overall outcome (and possibly on the individual-stage outcomes, depending on what is observable by the constituency).

Assume that the constituency establishes two criterion levels for values of X that will result in the executive being removed: one for the case of a first-stage conflict initiation and one for the case of a first-stage inaction. An executive decides (correctly) to initiate conflict but is unlucky and has an unfavorable outcome that would result in removal from office unless improved. At this point, removal may be all but inevitable if the executive makes peace because the losses already pass the removal threshold, and if peace has a low variance, the chance of raising the outcome sufficiently high to avoid loss of office is quite small. The alternative of escalation is attractive because, although the expected value of the outcome may be lower than that of peace, the greater variance of conflict holds out a better possibility of rising above the threshold for staying in office. This phenomenon that emerges from our model is sometimes called gambling for resurrection.

The main problem here is limited liability. Once the constituency seems determined to expel the executive from office for poor performance, it has no other sanction to apply. An executive, then, risks little in the way of further punishment by gambling and may gain the right to stay in office if the gamble is successful.

It is important to appreciate the difference between the gambling for resurrection rationale for executive behavior in deteriorating situations such as the United States experience in Vietnam and standard explanations such as George Ball's slippery slope. "Once we suffer large casualties, we will have started a well-nigh irreversible process. Our involvement will be so great that we cannot—without national humiliation—stop short of achieving our complete objectives" (Staw and Ross 1989, 216).

Note that Ball envisions an entire nation struggling to avoid "national humiliation." The principal-agent approach encourages us to view unprofitable escalation as a strategy used by a head of state to avoid the punishment costs associated with a poor outcome. Ball's slippery slope exists, but it is a personal one whose contours are designed to avoid the loss of political power. The two explanations also lead to sharply different predictions about the behavior of the successor president(s). According to Ball's model, we would expect any successor president who was elected before victory was ultimately achieved to be equally devoted to carrying out the war. Voters driven by national pride would demand it. Our principal-agent model would predict that the voters would do everything possible to elect a president who was committed to withdrawal.

In general, trying to distinguish gambling for resurrection from behavior that stems from nothing more than an executive carrying out the wishes of his or her constituency is no easy matter, but the modern history of unsuccessful interventions suggests that the agency explanation is important. The volume by Levite, Jentleson, and Berman (1992)

contains case studies of the most prominent unsuccessful interventions of the past three decades. In the case of each democracy or quasi-democracy (e.g., the United States in Vietnam, India in Sri Lanka, Israel in Lebanon, and South Africa in Angola), the government either continued to intervene or escalated the intensity of the conflict in the face of strong public opposition, just as the principal-agent model predicts.⁹ Vietnam may be the archetypical example of this, but one emerges from detailed descriptions of the situations in Lebanon, Sri Lanka, and Angola with the sense that, in each case, the intervention continued after the median voter had come to think that disengagement was the more sensible course. Moreover, a major part of the mandate of subsequent executives who took office after public antipathy to intervention reached its peak involved disengagement. Both of these occurrences are consistent with the principal-agent model; neither is what one would expect to see if an entire population was attempting to avoid national humiliation.

Gambling for resurrection as a motive for escalatory behavior is well-known but its connection to solutions for deterring undesired behaviors is not. Ellsberg (1972), for example, argued that the United States remained in Vietnam because no U.S. president could survive withdrawal, without ever suggesting that the organizational demands of uncertainty and the bad luck “trap” were responsible for this tragic state of affairs. The failure to connect the two processes is, in part, responsible for Ellsberg’s belief that a president could survive Vietnam by simply postponing total withdrawal. The principal-agent model tells us that this will not work. Once casualties reach a certain threshold, only victory, not the mere avoidance of defeat, is required. The distinction is important. *Ceteris paribus*, our model predicts that Vietnam would have cost Johnson his reelection in 1968; Ellsberg’s model suggests that it would not have.

Conclusion

This paper has presented a formal treatment of how a constituency can try to ensure that a head of state will act according to its wishes with respect to foreign conflict. This is a challenging problem because the head of state can be anticipated to possess substantial information advantages. We have seen how strategies that constituencies can use to deter overly passive and overly aggressive executives have a number of side effects. One such side effect is that they will sometimes lead to the removal of “innocent” executives after an unlucky outcome of a decision as to whether to initiate conflict, even though the original decision may have

⁹The same public opposition existed to the Soviet Union’s intervention in Afghanistan, but it is less clear here who the constituency was.

been optimal from the point of view of the public. This form of injustice is the price that must be paid for attempting to deter executives who do not share the public's preferences from ignoring those preferences in favor of their own.

A second set of side effects concerns the phenomenon known as gambling for resurrection, in which a conflict that should be terminated from the public's perspective is continued because it provides the only reasonable chance that the head of state can avoid removal from office. Recent case studies of intervention behavior in Vietnam, Sri Lanka, and elsewhere suggest that gambling for resurrection behavior is common and that it is the by-product of a principal-agent strategy rather than a constituency strategy.

One of the most interesting implications of the present study involves the trade-off between removing executives from office to cope with the principal-agent problem and minimizing gambling for resurrection. The more a constituency tries to limit future executive aggressiveness by increasing the punishment for poor outcomes, the more it creates an incentive for an executive who has initiated an unsuccessful war or intervention to continue the conflict, even though it would be terminated by a fully informed constituency. Like the president of a failing savings and loan in the 1980s, the executive has nothing to lose by trying ever more risky strategies. Conversely, the more a constituency seeks to mitigate this problem of gambling for resurrection by being sympathetic to an executive's argument that a war has gone badly for reasons that could not have been foreseen, the more that constituency will be led into conflicts that it would not, if fully informed, have entered.

The cost of confusing the two situations can obviously be very high. If a state thinks that it is suffering the effects of adventurism, when it is actually suffering the effects of gambling for resurrection—a real possibility given the lack of familiarity with the dynamics of the latter—it may move to increase the probability of removal for unsuccessful campaigns. This will further increase the incentive for gambling for resurrection, and the cure will have made the real disease worse than it was. We cannot speak to how frequently this kind of problem misspecification takes place, but its possibility suggests the merit of developing solutions for the agency problem that do not increase the incentive for gambling for resurrection and vice versa. A free press is an example of the former; parliamentary democracy is an example of the latter, since it permits a more timely removal of an executive than do fixed elections.

Manuscript submitted 30 November 1992

Final manuscript received 9 September 1993

APPENDIX

Optimal Bayesian Policies

In the main body of the paper, we examined the case where the populace has no reliable prior distribution on the preferences α of the executive and the private information θ . In this appendix, we consider the case where the populace does have such prior opinions about likely values of the private information variable θ (and thus a prior on the outcome itself) and about the type α of the executive. Suppose that the constituency's prior estimate of the value of the private information to be observed by the executive is given by $\theta = \xi + \bar{\theta}\tau$, where ξ and τ are known constants (prior mean and standard deviation) and where the density $g(\cdot)$ of $\bar{\theta}$ has mean zero and variance one (e.g., standard normal). Then the observed outcome X has a density h which is a convolution of f and g as follows:

$$h(x|\xi, \tau, \mu_{\theta I}, \sigma_{\theta I}, I) = \int \sigma_I^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta I}}{\sigma_I}\right) g\left(\frac{\theta - \xi}{\tau}\right) d\theta. \quad (\text{A.1})$$

In the specific case in which F and G are standard normal and $\mu_{\theta I} = \pm\theta$, we have that x is normally distributed with mean ξ and variance $\sigma_I^2 + \tau^2$. Contrast this with the executive's prior on x after observing θ , which is normal with mean θ and variance σ_I^2 . The executive's private information thus gives both a different estimate of the likely outcome and reduces the uncertainty, compared to the public's estimates.

Now the outcome that is observed depends on the executive's decision, which in turn depends on θ and α . First, consider the public-mirroring executive in which $\alpha = 0$. Then the decision to initiate conflict will be made exactly when $\theta > 0$. Consequently, the expected payoff to the public is

$$\begin{aligned} P_0 = & \int_{-\infty}^0 \int_{-\infty}^{\infty} x \sigma_0^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 0}}{\sigma_0}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\theta \\ & + \int_0^{\infty} \int_{-\infty}^{\infty} x \sigma_1^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 1}}{\sigma_1}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\theta. \end{aligned} \quad (\text{A.2})$$

If $\alpha \neq 0$, then the value of θ that divides decisions to initiate conflict from decisions to abstain is not zero. Let $t_0(\alpha)$ be that value of θ satisfying

$$E(u(x; \alpha) | \theta = t_0(\alpha), I = 1) = E(u(x; \alpha) | \theta = t_0(\alpha), I = 0). \quad (\text{A.3})$$

Then the payoff to the public from this executive's decision (with no sanction) is

$$\begin{aligned} P_\alpha = & \int_{-\infty}^{t_0(\alpha)} \int_{-\infty}^{\infty} x \sigma_0^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 0}}{\sigma_0}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\theta \\ & + \int_{t_0(\alpha)}^{\infty} \int_{-\infty}^{\infty} x \sigma_1^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 1}}{\sigma_1}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\theta, \end{aligned} \quad (\text{A.4})$$

which differs from the outcome of the ideal executive by

$$\begin{aligned} P_\alpha = & \int_{t_0(\alpha)}^0 \int_{-\infty}^{\infty} x \sigma_0^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 0}}{\sigma_0}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\theta \\ & - \int_{t_0(\alpha)}^0 \int_{-\infty}^{\infty} x \sigma_1^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 1}}{\sigma_1}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\theta \\ = & \int_{t_0(\alpha)}^0 \int_{-\infty}^{\infty} x \left(\sigma_0^{-1} f\left(\frac{x - \mu_{\theta 0}}{\sigma_0}\right) - \sigma_1^{-1} f\left(\frac{x - \mu_{\theta 1}}{\sigma_0}\right) \right) \tau^{-1} g\left(\frac{\theta - \xi}{\tau}\right) dx d\theta, \end{aligned} \quad (\text{A.5})$$

which is the cost of an executive that does not perfectly reflect the preferences of the constituency.

When a sanction is applied at z_p , the indifference point $t_Q(\alpha, p)$ of the executive is defined by the condition

$$\begin{aligned} E(u(X; \alpha) | \theta = t_Q(\alpha, p), I = 1) - QF(E_1 | \theta = t_Q(\alpha, p), I = 1) \\ = E(u(X; \alpha) | \theta = t_Q(\alpha, p), I = 0) - QF(E_0 | \theta = t_Q(\alpha, p), I = 0). \end{aligned} \quad (\text{A.6})$$

The payoff to the public from this option is

$$\begin{aligned} P_\alpha^Q = \int_{-\infty}^{t_Q(\alpha, p)} \int_{-\infty}^{\infty} x \sigma_0^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 0}}{\sigma_0}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\theta \\ + \int_{t_Q(\alpha, p)}^{\infty} \int_{-\infty}^{\infty} x \sigma_1^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 1}}{\sigma_1}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\theta \end{aligned} \quad (\text{A.7})$$

for fixed α and is

$$\begin{aligned} P^Q = \int \int_{-\infty}^{t_Q(\alpha, p)} \int_{-\infty}^{\infty} x \sigma_0^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 0}}{\sigma_0}\right) g\left(\frac{\theta - \xi}{\tau}\right) A(\alpha) dx d\theta d\alpha \\ + \int \int_{t_Q(\alpha, p)}^{\infty} \int_{-\infty}^{\infty} x \sigma_1^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 1}}{\sigma_1}\right) g\left(\frac{\theta - \xi}{\tau}\right) A(\alpha) dx d\theta d\alpha \end{aligned} \quad (\text{A.8})$$

when averaged over a prior distribution $A(\alpha)$ of α .

Now the first-order condition for optimal choice of p is

$$\begin{aligned} 0 = \int \frac{\partial t_Q(\alpha, p)}{\partial p} A(\alpha) \int_{-\infty}^{\infty} x \sigma_0^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 0}}{\sigma_0}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\alpha \\ - \int \frac{\partial t_Q(\alpha, p)}{\partial p} A(\alpha) \int_{-\infty}^{\infty} x \sigma_1^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 1}}{\sigma_1}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\alpha \end{aligned} \quad (\text{A.9})$$

where $\partial t_Q(\alpha, p)/\partial p$ is determined by implicit differentiation of (A.6) (note that E_0 and E_1 depend on p as does $t_Q(\alpha, p)$). The second-order condition is

$$\begin{aligned} 0 < \int \frac{\partial^2 t_Q(\alpha, p)}{\partial p^2} A(\alpha) \int_{-\infty}^{\infty} x \sigma_0^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 0}}{\sigma_0}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\alpha \\ - \int \frac{\partial^2 t_Q(\alpha, p)}{\partial p^2} A(\alpha) \int_{-\infty}^{\infty} x \sigma_1^{-1} \tau^{-1} f\left(\frac{x - \mu_{\theta 1}}{\sigma_1}\right) g\left(\frac{\theta - \xi}{\tau}\right) dx d\alpha \end{aligned} \quad (\text{A.10})$$

This strategy would clearly be more difficult to implement than the one described in the main body of the paper. It does, however, provide a yardstick against which more feasible strategies can be measured.

REFERENCES

- Alesina, Alberto, John Londregan, and Howard Rosenthal. 1993. "A Model of the Political Economy of the United States." *American Political Science Review* 87:12–33.
- Bueno de Mesquita, Bruce, Randolph Siverson, and Gary Woller. 1993. "War and the Fate of Regimes: A Comparative Analysis." *American Political Science Review* 86:638–46.
- Ellsberg, Daniel. 1972. *Papers on the War*. New York: Simon and Schuster.
- Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge: MIT Press.

- Gaubatz, Kurt Taylor. 1991. "Election Cycles and War." *Journal of Conflict Resolution* 35:212–44.
- Hibbs, Douglas. 1987. *The American Political Economy: Macroeconomics and Electoral Politics in the United States*. Cambridge: Harvard University Press.
- Kennan, George. 1951. *American Diplomacy, 1900–1950*. Chicago: University of Chicago Press.
- Levinthal, Daniel. 1988. "A Survey of Agency Models of Organizations." *Journal of Economic Behavior and Organization* 9:153–85.
- Levite, Ariel, Bruce Jentleson, and Larry Berman. 1992. *Protracted Military Interventions: From Commitment to Disengagement*. New York: Columbia University Press.
- Lippman, Walter. 1955. *Essays in the Public Philosophy*. Boston: Little, Brown.
- Morrow, James. 1991. "Electoral and Congressional Incentives and Arms Control." *Journal of Conflict Resolution* 35:245–65.
- Norpoth, Helmut, Michael Lewis-Beck, and Jean-Dominique Lafay. 1991. *Economics and Politics: The Calculus of Support*. Ann Arbor: University of Michigan Press.
- Persson, Torsten, and Guido Tabellini. 1990. *Macroeconomic Policy, Credibility, and Politics*. New York: Harwood Academic.
- Richards, Diana, T. Clifton Morgan, Rick Wilson, Valerie Schwebach, and Garry Young. 1993. "Good Times, Bad Times, and the Diversionary Use of Force." *Journal of Conflict Resolution* 37:504–36.
- Rogoff, Kenneth, and Anne Sibert. 1988. "Elections and Macroeconomic Policy Cycles." *Review of Economic Studies* 55:1–16.
- Staw, Barry M., and Jerry Ross. 1989. "Understanding Behavior in Escalation Situations." *Science* 246:216–20.
- Tirole, Jean. 1990. *The Theory of Industrial Organization*. Cambridge: MIT Press.
- Tocqueville, Alexis de. 1969. *Democracy in America*, trans. George Lawrence. 1850; rpt. New York: Anchor Books.