

The Informal Agency Theory

In Chapter 2, I argued that Huntington's model does not adequately explain American civil-military relations during the Cold War and should give way to a new theory. In this chapter, I draw upon the principal-agent framework to derive such a theory and explore its application to the civil-military problematique. I will discuss here the general features of the principal-agent framework and apply them to civil-military relations—how civilian and military actors have divergent preferences and private information relevant to the relationship; how these conditions complicate the way civilians and the military interact, and advance the possibility that the military will not work as civilians intend; how civilians can mitigate these problems with monitoring and punishment mechanisms. The informal discussion of agency theory in this chapter lays the groundwork for Chapter 4, in which I build a simple formal game of civil-military interaction.

Civil-Military Relations as Principal-Agent Relations

Civilians invent the military, contracting with it to protect society from enemies,¹ but then civilians find it necessary to assure themselves that the military will behave as intended. Relations between civilians and the military are, in their most basic form, a strategic interaction carried out within a hierarchical setting. It is strategic interaction because the choices civilians make are contingent on their expectations of what the military is likely to do, and vice versa. It is hierarchical (at least in democracies) because civilians enjoy the privileged position; civilians have legitimate authority over the military, whatever their de facto ability to control the military may be.

These two features—strategic interaction and hierarchy—are the distinctive features of the principal-agent framework, an approach developed by

economists to analyze problems of agency, where one person has delegated authority to someone else to act on his behalf (Alchian and Demsetz 1972, Niskanen 1971, Ross 1973, Bendor 1988). The employer (principal) would like to hire a diligent worker (agent), and, once hired, would like to be certain that the employee is doing what he is supposed to be doing (working) and not doing something else (shirking). The employee, of course, would like to be hired and so has an incentive to appear more diligent during the interview than he really is; this fact complicates the employer's efforts to pick the sort of employee who will want to work hard, a phenomenon referred to as the adverse selection problem. Once hired, moreover, the employee has an incentive to do as little work as he can get away with, all the while sending information back to the employer that suggests he is performing at an acceptable level; this fact complicates the employer's efforts to keep tabs on the employee and is called the moral hazard problem. The principal-agent approach, then, analyzes how the principal can shape the relationship so as to ensure that his employees are carrying out his wishes in the face of the adverse selection and moral hazard problems that attend any agency situation.

The most fruitful political applications of principal-agent have been in examinations of the way Congress or the president, or both, interact with the governmental bureaucracy.² The principal-agent framework has been used to explore whether the act of delegating policy implementation and, in some cases, policymaking power to bureaus and agencies is tantamount to an abdication of political control. I consider the basic concept of principal-agency as a "framework" rather than a "theory," since it is not a system of statements of cause and effect. But the framework can be adapted to yield a theory, which I do in this chapter and the next. To distinguish my application from the broader literature, I call my theory "agency theory."

The primary claim of the principal-agent literature is that delegation need not be an abdication of responsibility.³ Debate continues as to whether Congress or the president is the most influential in exercising political control, but the general consensus is that political control of the bureaucracy is greater than the problems of delegation would indicate. Politicians have remained dominant in policymaking even in arenas where they have apparently delegated responsibility to strong bureaucracies, as in the case of environmental regulation. The reason, the principal-agent framework suggests, is that principals are able to exert control in nonobvious ways, for instance by building incentives for good behavior into the contract or by establishing

third-party monitoring mechanisms that keep tabs on the agent for the principal. Even the absence of the more visible manifestations of oversight—intrusive audits and close monitoring of the agent's day-to-day behavior—does not mean the absence of control.

While most political science applications of the framework suggest the dominance of the principal (Congress) vis-à-vis the agent (bureaus), the conclusion that the principal solves all the problems of agency is not a necessary outcome of the model. Indeed, the framework lends itself to analyses of both the problems of agency and the possibilities for control.⁴

There are two prominent strands of analysis within this literature. One argues that agents work when monitored and shirk when not monitored, and so the solution to problems of agency lies in devising the optimal monitoring scheme, using intrusive and nonintrusive means (Kiewiet and McCubbins 1991; McNollgast 1987, 1989, 1990a, and 1990b; McCubbins and Schwartz 1984). Another strand argues that monitoring is inherently inefficient and that optimal compliance comes from improving the quality of the agent and bringing the agent's preferences more closely in line with those of the principal (Brehm and Gates 1997).

My own application of the principal-agent framework seeks to blend elements of both strands, incorporating considerations of how agents are monitored and also the extent to which the preferences of principals and agents converge. I also add a further consideration that is surprisingly absent from existing principal-agent treatments: how agents behavior is a function of their expectation that they will be punished if their failure to work is discovered; traditional principal-agent treatments assume punishment is automatic but, as I argue below, that assumption must be relaxed when analyzing civil-military relations. My agency theory, then, draws on, modifies, and contributes back to the general principal-agent literature.

There have been relatively few applications of this framework to civil-military relations. Deborah Avant (1993 and 1994) was the first to use it, borrowing insights from the principal-agent framework to explain different propensities for innovation across British and American military organizations. Risa Brooks (2000) also uses the framework to compare how different patterns of civil-military relations produce different grand strategies. Amy Zegart (1999) uses it to explore the design of national security agencies at the start of the Cold War. And Sharon Weiner (1997) likewise uses it to explore the motivations behind the Goldwater-Nichols reforms. None of these analyses attempts to derive a general theory of civil-military relations, the

project in view here, but they demonstrate the general utility of the approach; where appropriate, I note below how agency theory differs from these other analyses and why.⁵

In the civil-military context, the civilian principal contracts with the military agent to develop the ability to use force in defense of the civilian's interests.⁶ Once the contract is established, the civilian principal seeks to ensure that the military agent does what civilians want while minimizing the dangers associated with a delegation of power.

The process might resemble the following stylized narrative.⁷ Civilians recognize the need for instruments of violence, so they establish the military institution and contract with it the mission of using force on society's behalf. The contract is ritualized in the officer's oath of allegiance and reinscribed through a myriad of cultural symbols, such as the privileged place assigned to the military in celebrations of national holidays. The responsibility of the civilian does not end with this delegation of the protection mission.⁸ Civilians must decide what ancillary mechanisms they will establish to make sure that the delegation is not abused. These mechanisms have associated costs, however, making the monitoring decision tricky. Civilians could delegate authority with one hand but then institute such restrictive controls with the other as to effectively, if unintentionally, undo the delegation; conversely, the monitoring could be so lax that the military agent is essentially free to act as it wishes. The optimal mix of monitoring mechanisms is the one that minimizes the incentives and opportunity for the agent to flout the principal's wishes, at the least cost to the principal and while preserving the efficiencies of specialization that come with delegation.

The military may share the civilians' desire not to lose on the battlefield, but it would also prefer not to be subject to interference (which it might deem "meddlesome") by civilian authorities. Moreover, as we shall see, it may not share identical preferences with the civilians on all policy questions and so may seek to manipulate the relationship so as to prevail in policy disputes. In short, the military has the ability and sometimes also the incentive to respond strategically to civilian delegation and control decisions—in the jargon of principal-agency, to shirk rather than to work. (I realize the terms *work* and *shirk* carry problematic connotations in the civil-military context. As I explain at length below, the terms as I use them have a specific meaning and I use them only because all other synonyms are equally if not more ill-suited.) But civilians retain the ability to punish shirking if they discover that it is going on. Thus, the military decision whether to work or shirk is

shaped by how negatively those in the military view what civilians are asking them to do (how strongly, in other words, they would prefer to do it their way) and their expectation of the likelihood and severity of any punishment that might come their way should they shirk.

The process is iterative over time. As the external environment changes, for instance as threats to the state emerge or disappear, the civilians must revisit the problematique and make changes in the monitoring profile as needed. Of course, it is rather idealized to imagine civilians revisiting the civil-military problematique and deciding anew the question of delegation on every single security issue. In practice, the costs of reinventing the wheel probably make bureaucratic inertia (continuing the current pattern) attractive.⁹ Civilians still have the option, however, of changing the relationship on any issue; that they often choose not to simply underscores how they are sensitive to costs, as the agency model captures.

In sum, civil-military relations is a game of strategic interaction. The “players” are civilian leaders and military agents.¹⁰ Each makes “moves” based on its own preferences for outcomes and its expectations of how the other side is likely to act. The game is influenced by exogenous factors, for instance the intensity of the external threat facing the state made up of the players. The game is also influenced by uncertainties. The civilians cannot be sure that the military will do what they want; the military agents cannot be sure that the civilians will catch and punish them if they misbehave.

This rationalist approach to civil-military relations has its limitations. Some may find it off-putting to imagine the military choosing whether to obey or not based on crass calculations of self-interest. Indeed, most members of the military have a substantial moral commitment to what they do, and thus are motivated by more than resources or bureaucratic autonomy. The principal-agent approach does not rule out such nonmaterial incentives. On the contrary, the model establishes a rationalist baseline against which to measure the influence of these other considerations. It is, I would argue, only a point of departure, but a necessary one that has been undeveloped in the existing civil-military literature.

Divergent Preferences and the Work-Shirk Dimension

Like any agency relationship, the civilian delegation to the military could degrade into any number of suboptimal arrangements. The worst and most obvious are the central concerns of the civil-military problematique: col-

lapse on the battlefield or military overthrow of the regime. Other less dramatic but nevertheless unsavory options are possible. The military could attempt to drag civilian leaders into an unwanted war, which, even if successful, would impose costs they might otherwise choose not to pay. The military could on a sustained basis extract higher resources from society than are really necessary. The military could successfully resist civilian dictates, perhaps by claiming that the action commanded is impossible (or too costly) or simply by failing to carry out the order. In this analysis, I focus attention on one set of conduct that I call shirking, when the military does not work as civilians direct.¹¹ This set is of particular interest for democratic civil-military relations because, by definition, civilian preferences must prevail over military preferences in a democracy.

The colloquial meanings of work and shirk are not particularly helpful, and the reason points to an important limitation of economic models in political analysis. The problem in civil-military relations is not a lazy military—or at least this is not the only nor the most important problem. While individual military officers might be lazy or shiftless, the general problem is more complex than simply keeping the military industrious.¹² In economic relationships, the principal and the agent have a different set of incentives regarding the basic work assigned to the agent: the principal wants lots of work for little pay, and the agent wants lots of pay for little work. In economic settings, it is not implausible to imagine that an agent would produce no work if he could get away with it. In the civil-military setting, this rarely makes sense. It is more reasonable to posit that both the civilian principals and the military agents want the same thing: security for the state. They can, however, disagree on how to provide that security, in general and especially in particular settings.

Areas of disagreement wax and wane—the preferences of civilians and the military can diverge and converge. As explained in greater detail below, one way civilians can shape military behavior is to seek a convergence in views by promoting military agents who hold preferences more similar to those of civilian principals. But there is a limit to how far convergence can go; perfect overlap is not feasible. For starters, military communities have strong identities that mark them as “different” from those of civilians, and this is deliberately cultivated and signified through uniforms, oaths of office, rituals, and so on; there is, in other words, some irreducible difference between military and civilian, and this will naturally extend to different perspectives. Moreover, the civil-military difference is compounded by the dif-

ferent role each plays, one as principal, the other as agent; there is a *de minimis* difference in perspective that attends agency, hiring someone else to do something for you. The empirical record bears out this fact: there are undoubtedly instances when civilian and military preferences greatly converged in the United States, but these instances have been limited in time and space and the preferences have not overlapped perfectly throughout the period under study. In sum, the agency problem of working and shirking arises because of civilian-military disagreement over means, if not ends, which itself arises from inherent differences in the roles played by civilians and the military.

Working is doing things the way civilians want, and shirking is doing things the way those in the military want. Shirking, in my use, has an explicit civil-military context. Shirking is part of a broader range of deviant behavior in which a soldier might engage—for instance, looting, going to sleep on duty, showing insubordination to an officer, mistreating prisoners of war, or failing to clean one's weapon (Bryant 1979). Of course, the civilian wants the military to obey all laws—not to rape, murder, steal, and so on. Such crimes as these rarely engage the civil-military relationship per se, although they can develop a civil-military connection, as the so-called “sex crimes” that preoccupied public attention in the spring of 1997 demonstrated.

Sometimes the behavior is readily categorized: a coup, General LaVelle's unauthorized bombing during the Vietnam War, and the My Lai massacre would all be clear cases in which civilian will was flouted. Other behavior, however, is far more difficult to pin down: is the military shirking when military estimates of what a given mission will cost are higher than civilian estimates? Or when the military-dominated review of service roles and missions returned the “finding” that there were no savings to be gained by consolidating the services, even though the civilians who asked for the review clearly thought there would be? (See Gordon 1992a, Schmitt 1993e.) The agency framework could consider these kinds of conduct to be shirking, even though they are obviously more ambiguous and less drastic than a coup.

Working, in the broadest sense of the word, means doing something to the principal's satisfaction. Shirking means not doing it to the principal's satisfaction. Since there is always the possibility of unforeseen circumstances beyond the control of either principal or agent that can stymie the agent's work, most principal-agent treatments narrow the meaning of work and shirk further: working involves a good faith effort to represent the princi-

pal's interests, or put another way, working is the ideal conduct that the agent would perform *if* the principal had full knowledge of what the agent could do and was in fact doing.

The issues at stake in civil-military relations, however, introduce a complexity that is missing from economic applications of agency. Working, and hence its opposite, shirking, are multidimensional because civilian desiderata are themselves multidimensional. Recall the problematique discussed in the first chapter. Civilians want protection from external enemies *and* want to remain in political control over their destiny. The first goal may be called *functional* and the second may be termed *relational*. The agent may act in ways contrary to either the functional or the relational goal. Both inimical activities must be considered shirking.

Both the functional and the relational goal may be further disaggregated into tasks. The functional goal includes the following:

1. whether the military is doing what civilians asked it to do, to include instances when civilians have expressed a preference on both the "what" and the "how" of any given action;¹³
2. whether the military is working to the fullest extent of its duty to do what the civilians asked it to do;¹⁴
3. whether the military is competent (measured by some reasonableness standard) to do what civilians asked it to do.

The relational goal can be broken down into the following:

1. whether the civilian is the one who is making key policy decisions (i.e., no de facto or de jure coup) and whether those decisions are substantive rather than nominal;
2. whether the civilian is the one who decides which decisions civilians should make and which decisions can be left to the military;¹⁵
3. whether the military is avoiding any behavior that undermines civilian supremacy in the long run even if it is fulfilling civilian functional orders.

The functional component to shirking may be more obvious, but the relational component is at least as important and often more difficult to secure. For instance, the military may appear to obey civilian orders (it does everything the civilian asks it to), but that is only because the military is so powerful that the range of issues on which civilians can in fact ask it to do something is very constrained. If civilians are not able to set the boundaries for

which decisions they choose to delegate to the military, then the relational goal is not met and the military is, *de facto*, shirking. As Welch (1976, p. 317) observed: “Perhaps the best measure of the strength and extent of civilian control of the military is governmental ability to alter the armed forces’ responsibilities.”

When the answers to all the functional and relational component questions are positive, then the military is working. When the answers are negative, then the military is shirking. Obviously, the responses can be mixed, yielding a gradation of working and shirking. At the extreme end of shirking is the traditional civil-military concern of a coup. At the extreme end of working is some ideal-type military that does everything the civilian has contracted with it to do, vigorously and without subversion.

The military’s advisory role further complicates the concept of shirking in the civil-military context. Working does not imply that the military immediately and mutely executes every harebrained scheme that issues from the mouth of any civilian policymaker. Part of the military obligation is to advise civilians on the military implications of proposed courses of action. As Chapters 5, 6, and 7 will demonstrate, there is an exceedingly blurry line between advising against a course of action and resisting civilian efforts to pursue that course of action. Sometimes negative advice can rise to the level of shirking, especially if the advice is exaggerated. As one traditional treatment of civil-military relations put it, there is a difference between offering advice and “insisting that it is absolutely necessary to pursue a certain course of action if disaster is to be avoided”—where the latter might reach the level of shirking (Sapin and Snyder 1954, p. 55). Nevertheless, shirking is not synonymous with “persuading a civilian policymaker to change his mind.” Thus, evaluating whether shirking has occurred is not as simple as discovering whether military advice was followed. Rather, it involves judgments about the integrity of the military advice itself as well as judgments about the conditions under which civilians changed their minds. Were military advisors exaggerating (or minimizing) the costs of a course of action so as to tie the hands of the policymaker? Did the policymaker abandon a course of action because military resistance was too strong, or because he was truly persuaded that it was unwise?

Just as civilian preferences are multidimensional, so too are the preferences of the military agent. A weakness in early applications of the principal-agent framework to political problems is that they overemphasized the perspective of the principal (Hamilton and Schroeder 1994, Moe 1987,

Spence no date). A simple-minded translation of the economic principal-agent relationship to the political realm misses an important aspect of the agent's motivations. Whereas economic agents may generally prefer shirking to working, political agents are likely to be motivated (at least in part) by a substantive interest in the policy itself. In short, there is a reason why an agent ends up working for the Department of Defense rather than the Department of Health and Human Services, and this has important implications for how the agent reacts to the principal (Brehm and Gates 1997). Specifically, the military agent is assumed to have three sets of preferences: over policy outcomes, over how his behavior is interpreted, and over how the relationship is monitored.

First, unlike an economic agent who might not care how many widgets he produces, the military agent cares about policy and has a general idea about what should be done.¹⁶ The military agent would like to be told to pursue the policy he wants to pursue. While what the military agent wants will vary from situation to situation, the traditional civil-military relations literature has identified several standard features. The military agent is willing to risk life, but would prefer not to die needlessly (Huntington 1957, pp. 68–69).¹⁷ The military agent would prefer to deal with threats from a position of advantage, controlling the tempo and the scope of the conflict. This translates into a general preference for offensive operations and even preventive operations, dealing with problems before they become unmanageable (Posen 1984, pp. 47–49; Sagan 1994, pp. 75–76). Moreover, the foregoing logic also suggests that the military agent would be more likely to inflate threats and inflate requirements for meeting those threats so there would be less chance of being taken by surprise (Allison 1971).

Related to policy preferences, but nevertheless distinct, is the second military preference, what might be called a general military preference for honor, or a desire for respect.¹⁸ This is not to be confused with glory, or a desire for distinction, which connotes the militarism caricature. Rather, honor here captures the Aristotelian idea of getting credit for doing what is right, where right is acknowledged by others and defined according to some generalized conception of the good (*Nichomachean Ethics*, Book IV, No. 3). The desire for honor derives partly from the basic human desire for legitimacy—the desire to gain peer approval. It is worth noting that even the most heinous scoundrels in history sought to justify their actions, however lamely and self-servingly. Military men and women, like all men and women, partly seek the approbation of others, especially peers. Honor has always

played a central role in military life.¹⁹ Honor permeates the famous concept of small-group cohesion, the factor that makes human beings willing to risk their lives (Holmes 1985, pp. 290–315). In the civil-military relationship, the military preference for honor can be used to reinscribe the principle of civilian control; the military subordinates itself to civilians because, in a democracy, such subordination is recognized to be right and it would dishonorable to do otherwise (Janowitz 1971, 439–440). In democracies, civilian governments are presumed to be legitimate, and this general presumption is shared even by the military. Thus the preference for honor can work to mute the impulse to shirk for military agents, even when other factors (as discussed below) indicate they should. Traditional civil-military relations theory relies extensively on honor (also called the “ethic of subordination” and “professionalism”) to explain civilian control. It can be incorporated in a principal-agent framework, but it is recognized as just one of the factors shaping the military’s preferences.

The third basic military preference concerns the monitoring relationship. Regardless of what the military agent is asked to do, he would like to do it with the minimum of civilian interference and oversight. As in traditional organization theory, the military agent prizes autonomy—policy autonomy, the ability to decide what to do, and implementation autonomy, the ability to decide how to do it (Betts 1991, pp. 5–15; Posen 1984, pp. 52–54; Lebow 1987, pp. 76–79; Bouchard 1991, p. 30).

These three sets of preferences set up the possibility that the military might do what it wants instead of what the civilian wants—in other words, that it might shirk instead of work. If the military shirks, of course, it is evidence that the agency relationship has broken down, at least in that instance. It is important to note, however, that pure working does not necessarily mean that the outcomes of military action will please the civilian. This is because the functional goal itself, security, involves another actor not in the principal-agent relationship: the enemy. The functional goal of security can be broken down into tasks—develop a military capable of winning two medium-size regional conflicts nearly simultaneously, integrate women into combat roles, invade Haiti, and so on—but faithfully fulfilling those tasks may or may not provide ultimate security for the country. Indeed, some of the things the civilian wants done may in fact work contrary to the overall goal of providing security for the state. Moreover, in fulfilling the principal’s functional directions, things can go wrong simply due to Murphy’s Law or Clausewitzian “friction.” Thus, working and shirking are not synonymous

with winning and losing on the battlefield. Battlefield victory is a result of a two-sided interaction. One side can “work” and still lose if the other side has superior forces, or if the objective is ill-conceived. Likewise, one side can “shirk” and still win if the other side is even less competent, or if the task the civilian wanted done (the work) was in fact not appropriate for the security goal.

The security goal can even be met under suboptimal circumstances that would displease the civilian if he had perfect information. First, consider the case when the protection afforded by the military is costing more than it needs to cost. Here the civilian achieves the functional goal but overpays for it. A less costly military capability would suffice. Second, consider the case when the protection afforded by the military is sufficient but involves running undesirable risks. Here the civilian achieves the functional goal, but just barely and while skirting near disasters that a different policy mix might avoid more comfortably. Knowing how much is really enough, even with twenty-twenty hindsight, is so intractable as to make this line of inquiry less fruitful than focusing on other, more readily ascertainable instances of shirking. Even if the action is not displeasing to the civilian *ex post*, it still counts as shirking if it involves the military’s taking a deliberate action that violates standing orders.²⁰

The foregoing suggests that civilians and the military can share the goal of national security (and in the U.S. case, one can even stipulate this as largely true), but also that *civilians and the military are both imperfect judges of what is needed for national security*. The principal-agent problem arises when there is disagreement over what is needed or appropriate for national security, whether or not one side is “correct” about what is in fact needed.

In a democracy, civilians have the right to be wrong. Civilian political leaders have the right to ask for things in the national security realm that are ultimately not conducive to good national security. The military should advise against such policies, but the military should not prevent those policies from being implemented. While this view is a necessary and logical conclusion of the premises of democratic theory, it is nonetheless controversial. Many observers, especially military observers, may be more sympathetic to S. L. A. Marshall’s view, as articulated in *Men against Fire*: “our Army should never be put under the necessity of humoring and yielding irretrievable ground to the inevitable minority of malcontents or of permitting governing principles to be influenced by voices from the lunatic fringe, even those which have been elected to Congress” (Marshall 1947, p. 165). Marshall’s

view may be appropriate if he is talking about *minority* positions within the civilian principal, but he misunderstands the military role if he advocates that the military ignore foolish “voices” that represent a governing civilian position. If the politicians are in fact wrong, then they are shirking in the deeper voter-as-ruler principal-agent relationship, and the voter-as-principal is obliged to punish the politician-as-agent by voting him or her out of office. It is arguable whether the military is in fact better able to judge the true national security needs of the polity; but even if true, in a democracy the military is not the one assigned to ensure that civilian politicians are not shirking.

This can be represented graphically as three points in a three-dimensional space, where each dimension represents a critical component of national security policy (for example, force structure, grand strategy, and operational plan). One point represents the policy mix that would produce true optimal security; this is what the civilian and the military ultimately want. Another point represents the civilian principal’s desired policy mix, which is the civilian’s best estimation of what is needed for security; this is what the civilian asks for. The third point represents the military agent’s desired policy mix; this is what the military asks for. The work-shirk continuum concerns only the nearness of behavior to the civilian or military desired point, and does not directly address whether the output approximates theoretically optimal security.²¹

Even if the distances between the civilian, the military, and the true point are trivially small, however, the principal-agent problem still arises because of the relational imperative. Civilians still have to be calling the shots, and military behavior has to be consistent with civilian supremacy. Shirking can arise even if the military is doing what the civilians have asked for, and even if what the civilians have asked for will produce what the civilians want. Even under these happy conditions, it still matters whether the civilians are the ones who make key policy decisions (no *de facto* or *de jure* coup), whether the civilians are the ones who decide which choices civilians should make and which can be left to the military, and whether the military is behaving in a way that supports civilian supremacy in the long run. If the answers to any of those questions are negative, then the military is shirking, even though what it wants is not significantly different from what civilians want.

Shirking can even arise in advance of civilian directives, if the military acts in such a way as to tie the hands of its civilian leaders. Perhaps the most im-

portant example of this kind of shirking came in the context of the U.S. Army's response to its experience in Vietnam. Senior army leaders believed that one of the lessons of Vietnam was that the army should never be deployed without the full support of the American public.²² Furthermore, the leadership believed that such support would come only if the civilian political leaders spent the political capital necessary to mobilize public opinion—and, finally, that civilian leaders would spend such political capital only if they were unable to deploy military forces any other way. In other words, the army believed that civilian leaders should not be given the opportunity to deploy combat troops on what appeared to be a “costless” mission; civilian leaders should commit to paying political costs, and those costs should be frontloaded. The army hardwired this lesson into its force structure, deliberately shifting to the reserve component key support functions, so that the army could not be deployed without mobilizing the reserves—precisely the politically costly move President Johnson had refused to take in the early phases of the Vietnam War. As one interviewer posed the question to a senior army general: “Was part of the thinking in integrating the reserves so deeply into the active force structure that we were making it very difficult, if not impossible, for the President to deploy any significant force without calling up the reserves?” The general replied, “That’s it, with malice aforethought” (Sorley 1992, pp. 363–364). The issue is not whether the army was correct in interpreting the Vietnam War, nor even whether it is wise for the president to mobilize public support for a combat operation before committing U.S. troops to such an operation. The army may be correct on the substance, but the effort to tie the hands of the president constituted a usurpation of the civilian leadership’s role in deciding when and how to use force.²³ Then Secretary of Defense Schlesinger claimed that the army was not being insubordinate, but contradicted his own assessment by observing that “the military sought to fix the incentives so that the civilians would act appropriately” (Sorley 1992, p. 364). This is, of course, precisely the principal’s function in a classic principal-agent relationship, and it is certainly not the role of the military-agent.

An important exception is the case in which the civilian principal, like Odysseus, asks the military agent to tie his hands in some way so that the civilian can get what he knows he ultimately wants and not what he will *say* he wants under some limited circumstances. A classic example is the Base Realignment and Closure Commission (BRAC) established by the U.S. Congress to identify which military bases to close in the post–Cold War

drawdown. Congress knew that political considerations would make selecting bases impossible on a case-by-case basis, so it tied its own hands and delegated the selection authority to a separate commission. When the BRAC posted the list of closures, the individual members could and did protest vigorously and seek desperate measures to undo the BRAC, even as Odysseus tore at his chains, but in context received the outcome for which they had contracted.

To sum up, shirking and working are multidimensional concepts, consisting of both functional and relational components and reflecting the multidimensional and possibly divergent preferences of the civilian principal and the military agent. The agent is said to work perfectly when it does what it has contracted with the principal to do, how the principal has asked it to, with due diligence and skill, and in such a way as to reinforce the principal's superior role in making the decisions and drawing the lines of any delegation. The military agent is said to shirk when, whether through laziness, insolence, or preventable incompetence, it deviates from its agreement with the civilians in order to pursue different preferences, for instance by not doing what the civilians have requested, or not in the way the civilians wanted, or in such a way as to undermine the ability of the civilians to make future decisions.

In practical terms, military shirking in the U.S. context is rarely open insubordination and has never risen to the point of a coup. But shirking is possible even if the military never carries out a coup, and when it happens, shirking by the U.S. military usually takes one of three forms: (1) efforts to determine the outcome of a policy calculus by giving inflated estimates of what a military operation would cost; (2) efforts to determine the outcome of a policy calculus with "end runs," unauthorized public protest, leaks, or appeals to other political actors; (3) efforts to undermine a policy through bureaucratic foot-dragging and "slow rolling" so that the undesired policy will never be implemented.

Information Asymmetries in Civil-Military Relations

Principal-agent relationships involve information asymmetries. Both sides share common information; in the civil-military context, they know who the domestic players are, the size of the defense budget, the general identity and nature of their enemies. They also share a common history and political memory. But each has private information that is discerned only dimly by

the other. The military agent's status as an expert on the management of violence confers significant informational advantages over civilians in areas like tactics and logistics. Of course, the extent to which the military agent's expertise exceeds that of the civilian principal, even in the arcana of operational art, varies with the backgrounds and résumés of the individuals involved; it is not uncommon for civilians on the staff of the Office of the Secretary of Defense to have more technical expertise than their military counterparts (Gibson and Snider 1997). But on average the military agent will have devoted more time (and more recently) to developing this technical expertise than will have his civilian political principal. Likewise, crucial aspects of military behavior and even military predilections may be unknown to civilians. For instance, the civilians cannot know for certain whether the military is inclined to shirk, nor can they know if the spirit of their orders is being carried out. Moreover, as the operation of the military moves closer to combat, civilians are at an even greater informational disadvantage. It is hard enough to monitor the activities of the military when it is bivouacked near the capital. When it is deployed on a distant battlefield in the fog of war, communications difficulties could render monitoring impossible for even the most attentive civilian leader.²⁴

Likewise, some information is private to the principal. For instance, only the principal knows exactly how he judges various risks and how these judgments translate into preferences over outcomes. While the civilian principal may convey this information to the military agent in the form of orders, it is also possible that exogenous changes in nature—the outcome of military operations, the arrival of new threats and national challengers, and so on—will cause a shift in the preferences of the principal. The anticipated effect of these changes will certainly be hidden from the military agent, and may even be hidden from the principal himself. In other words, there is sufficient information hidden from the military agent to warrant fear of a “stab in the back” from civilian leaders; will civilian leaders abandon the military when things go sour, the quintessential fear that constitutes the “Vietnam syndrome” among senior military officers? In general, however, information asymmetries favor the military agent.

The special concern at the heart of the civil-military relationship—controlling the use of deadly force—introduces further peculiar twists into the standard principal-agent scenario. The first and most obvious distinctive feature is that the stakes are much higher. If an elected representative comes a cropper, the damage to the polity (even with the highest elected office) is

bounded. Lousy political agents can commit many sins of omission and commission, but they are hard-pressed to bring down the republic. Failure to get the military agency problem right can result in one of two grave disasters: the military agent may turn on society and rob it of its political freedom, or the military agent may fail on the battlefield and leave the polity vulnerable to conquest by external enemies. In almost all cases, human lives will be lost in the process. Thus the cost of failure, the price of trial and error, borders on the prohibitive. Of course, the fate of the republic does not hang on every military issue and there are many mundane matters in which the stakes seem small. In general, however, the consequences of failure are profound and this will cast a shadow over the actors' decisionmaking.

It is also plausible that the basic information asymmetry problem inherent in any principal-agent relationship is particularly acute in the civil-military relationship. Traditional theory emphasizes the unique expertise of the military officer: the management of violence. Certainly civilians can gain expertise on a wide variety of defense policy issues, but civilians, by definition, leave combat to the military, and combat is the distinctive mark of military expertise.²⁵ Like many other complex policy issues, questions about technical competence and specialized knowledge exacerbate the basic informational challenges facing civilian principals.

Another distinctive feature is that the military is perhaps the only profession that never really gets a chance to practice. To be sure, the military can train and rehearse, but the essence of combat—wielding force against a determined enemy—cannot be simulated reliably. The unreliability of performance indicators is a common problem in principal-agent relationships, but the military condition adds an additional layer of uncertainty. Neither side, neither civilian nor military, can be sure about the military's type, at least with respect to performance on the battlefield. In conventional analyses, the expert agent generally understands the consequences of his actions even if the political principal may not. The military expert can claim that his expertise narrows somewhat the uncertainty boundary, but he still will not know for certain whether he will fail catastrophically in battle (Rogers 1940, p. 283). The military agent *will* have private knowledge about his inclinations to shirk, and will certainly know whether his day-to-day activities track with the performance indicators established by the civilian principal—neither of which the principal will know with certainty—but neither the military nor the civilians will know for sure what all this means for the ulti-

mate purpose of the military: the probability that the military can protect society when challenged in a war.²⁶

The information problem is further exacerbated by the secrecy restrictions that accompany military actions. A common obstacle in principal-agent relationships is the tendency of the agent to withhold information that reflects unfavorably on the behavior of the agent. The information classification system vastly eases the task of an agent who wishes to keep inconvenient information from being disseminated. While in theory the civilian principal may be entitled to know everything, in practice the costs of that are great in any principal-agent relationship. Those costs become prohibitive when secrecy laws, with strict punishments for the release of classified information, reinforce the agent's natural tendencies. Secrecy laws enjoy legitimacy because they are thought to contribute to the general protection of society. By association, this confers at least a certain amount of legitimacy on military efforts to keep principal civilians in the dark. Moreover, since so many of the principal-agent control mechanisms are essentially efforts to open the agent's hidden behavior to public scrutiny and thereby to alert the principal to improper agency, they are much less appropriate in contexts where even the principal agrees that the agent's behavior *ought* to avoid general public scrutiny (Lindsay 1994a, p. 283). Without this mechanism, however, it becomes easier for the agent to abuse the system, as was evidenced by Oliver North's off-budget covert operations.

Still another distinctive feature is the problem of competence. In its original microeconomic formulation, the principal was assumed to be as expert as the agent in all matters; the agent is hired merely to ease the work burden on the principal. A senior manager in a business has presumably worked his way up, learning the ropes of the junior positions, and hires workers merely for the sake of efficiency. Political science applications have tended to emphasize the difference in expertise between the principal and the agent, where the former may not be technically competent to do the things asked of the agent. Democratic theory further blurs the issue with the notion of political competence discussed earlier—the agent is not competent to judge risks even if he has special, even unique, technical competence (Dahl 1985). The special case of civil-military relations adds one further wrinkle: the military agent is asked to put his life on the line to protect the civilian. By virtue of his willingness to sacrifice, the military agent may be thought of as possessing a special moral competence, balancing somewhat the political com-

petence of the civilian principal. This belief in moral competence serves to muddy the lines of authority between civilian and military, particularly when the civilian is directing the military to put itself in harm's way. The *de jure* hierarchy may be unchanged, but the moral ambiguity of the relationship bolsters the hand of a military agent should he choose to resist civilian direction.

The relative moral competence of civilians and the military ebbs and flows both with the qualifications of idiosyncratic civilian leaders and with changes in the modal career paths of civilians. Thus, civilians like Dwight Eisenhower and Bob Dole, who served in the military, would bring a larger measure of moral competence than would civilians like Bill Clinton and Newt Gingrich, both of whom went to extraordinary lengths to avoid military service. One of the most intriguing trends in U.S. civil-military relations has been, first, the rise of military civilians during the Cold War—as civilian political leaders shared a military experience in World War II, Korea, Vietnam, or at least in peacetime service—and now the rise of purely civilian civilians (Bianco and Markham 2001, Gibson and Snider 1997, Feaver and Gelpi forthcoming).

In combination, these novel features have a profound effect on the principal-agent relationship. Because the stakes are so high, the civilians have an incentive to revisit and tinker with agreements and conditions they may have found acceptable before. Because the military cannot claim exclusive knowledge about at least one specific value at stake—the probability of success on the battlefield—the civilian has grounds to justify this meddling. Because of classification restrictions, the military has an extra advantage in withholding information, thereby frustrating civilian oversight and control. Because the military is prepared to do something that the civilian is not required to do, the military has a formidable moral arsenal with which to fight unwanted civilian interference. In short, the expectation is that this principal-agent relationship should be particularly characterized by distrust and friction, and any equilibria of delegation and control are unlikely to endure, giving way instead to new arrangements as costs and benefits shift.

Adverse Selection and Moral Hazard in Civil-Military Relations

The interaction of divergent preferences and informational asymmetries produces two problems: adverse selection and moral hazard. Adverse selec-

tion refers to the moment of hiring in the employer metaphor. Has the employer hired someone who is naturally a hard worker or has he been deceived by the interview and hired a lout? Just how closely aligned are the preferences of the agent and the principal? The adverse selection problem means, in the first instance, that the employer cannot know for certain about the true preferences and capabilities of the applicant. But adverse selection is more than mere uncertainty about the applicant. It also refers to the fact that the very act of hiring creates perverse incentives for the agent to misrepresent himself, which thereby increases the chances that the principal will hire a lout: it is hard to verify the true type, and the lout has a great incentive to appear even more attractive than a good worker.²⁷ Indeed, because the employer offers a wage that is pegged to attract someone with the statistically average set of qualifications needed for the job, the job will be especially attractive to louts, who will know that the offered wages are higher than their own true worth; considerably more diligent employees are likely to find any given job less appealing because they know that the average wage understates their true value. More generally, adverse selection can extend beyond the hiring phase to include all those situations in which the agent presents himself, or some proposal, to the principal for approval or decision. For instance, it means that because of their informational advantage over superiors, subordinates tend to propose policies that benefit their own interests rather than the interests of the superiors.

In a civil-military context, the adverse selection problem shows up in the accession of personnel into the military, in the promotion of individuals up the chain of command, and in the ongoing give and take of military policy. First, it is at least plausible that the peculiar mission of the military—to kill people and blow things up—attracts a special kind of person, one who may make the principal-agent relationship particularly problematic. It is reasonable to expect that the demanding mission would attract people with a sense of adventure, a tolerance for hardship, a commitment to order and discipline, and so on. It is possible, however, that these same qualities (or others correlated with them) accentuate the difference between civilians and the military and so lead the military to be especially distrustful of civilian leadership. One of the major concerns of traditional civil-military relations theory was precisely the great divergence of viewpoint between what Huntington called the liberal civilian ideology and the military mind (Huntington 1957, pp. 59–79). This does not mean that all military officers think alike. As we shall see, manipulating differences of opinion *within* the military is an im-

portant method of civilian control. Adverse selection could suggest that the civilians are liable to pick either people who are poor warriors (perhaps because they pick people like themselves) or good warriors who are likely to resent their authority. The decision to “hire” the military is revisited at regular intervals in decisions to promote or to fire certain individuals. Of course, as personnel proceed through the ranks the uncertainty should ease somewhat, since civilians gain more and more information on which to base such decisions.

Adverse selection also crops up in the uncertainty civilian leaders have in evaluating proposals originating from military organizations. Is this budget request necessary to accomplish the mission, or is it padded to serve the military organization’s interests? Again, because the military has an information advantage it can advance artfully drawn proposals that appear to meet civilian needs but in reality are tailored to its own interests. In the extreme, adverse selection might lead civilians to adopt policies they think will increase the military’s ability to protect society but that in fact will increase the ability or even the propensity of the military to undermine society.

Moral hazard refers to the behavior of the employee once hired. Like adverse selection, moral hazard refers at a general level to the problem that principals cannot completely observe the true behavior of the agent and so cannot be certain whether the agent is working or shirking. It has an additional specialized meaning based on the perverse incentives in the agency relationship.²⁸ Employees have an incentive to shirk rather than work; if you can get paid for doing less, why do more? The principal, of course, tries to minimize shirking because it is inefficient. The best way to minimize moral hazard is to reward (or punish) the agent based on whether he is working (or shirking). If the behavior of the agent is hard to monitor, however, how does the principal know whether the agent is really working or shirking? In many principal-agent relationships the behavior of the agent is hard to observe. In these cases, performance is usually measured by proxies that substitute, with some loss of validity, for the true goal (in an academic setting, however, number of publications substitutes for scholarly contribution to one’s field). Once established, however, the workers have an incentive to optimize on the indicator rather than on the true behavior desired. The gap between the indicator or rule and the desired behavior can be so great that devious subordinates can bring organizations grinding to a halt simply by strictly observing the official rules. Indeed, “work-to-rule” is an

effective form of workers' revolt in large complex bureaucracies like post offices and police forces.

Moral hazard pervades the civil-military relationship. How do we know that the military is doing what it is supposed to be doing? How do we know that the military is serving the interests of the country and not parochial interests, either of individual officers or of some larger group (such as a service or branch)? The problem is especially acute because the real goal of the military, being ready to protect society from its enemies, is not directly observable most of the time. In its stead, we can observe what the military is doing in terms of training, buying weapons, and so on. But how can civilians ensure that the peacetime behavior they are able to monitor correctly indicates that the military will perform as directed during wartime? Moreover, if in the meantime the military is rewarded based on a set of indicators that only imperfectly measure the true desired output, the military has a strong incentive to optimize on those indicators, not on the true output.

Mechanisms for Civilian Oversight of the Military

A central premise of political applications of the principal-agent framework is that despite all of the foregoing problems, political control does not end with the delegation decision. Civilians still have means available with which to direct the military and thereby mitigate the adverse selection and moral hazard problems inherent in delegation. In essence, control or monitoring mechanisms are ways of overcoming the information problems discussed above, perhaps by getting the agent to reveal information or perhaps by adjusting the incentives of the agent so that the principal can "know" that the agent wants what the principal wants.²⁹ The difficulty of observing battlefield operations does not in and of itself preclude close monitoring. The principal-agent framework cues us to look for operational control measures in nonoperational contexts. For instance, control over budget and doctrine could be surrogates for control over the performance of the military on the battlefield—an arena the principal cannot directly oversee. By shaping budget or doctrine in a certain way, the principal can know something about the likely activity of the agent, even without directly observing him.

Consider the most obvious form of monitoring: restricting the scope of delegation to the military. Some degree of agency is inevitable in modern civil-military relations. Not everyone can go to the battlefield to fight. But in

theory any amount of delegation short of that basic distinction is possible. Force management can be broken down into three broad categories: strategy, structure, and operations. These broad categories can be further broken down into still smaller discrete steps. Consider the hypothetical case of a single use of force, say the decision to bomb a Serbian artillery site outside Sarajevo. This might consist of: the decision to flip the switch that drops the bomb; the decision to use a manned aircraft to deliver the ordnance; the decision to target that particular artillery piece; the decision to target an artillery site; the decision to use force against the Serbs; the decision to establish an artillery exclusion zone; the decision to defend the Muslim enclave at Sarajevo; the decision to get involved in the former Yugoslavia; the decision to commit to security and stability in Central Europe; the decision to commit to NATO; the decision that Europe is a vital national interest; and, finally, the decision to be engaged in world affairs rather than isolationist. One could imagine another parallel or interwoven series of decisions that would walk back from the ordnance and aircraft, through the pilot's training regimen, through the aircraft production, to line items in the defense budget.

Civilians could devise the strategy, deduce operations and battle plans therefrom, specify tactics to achieve those aims, outline logistics and equipment needed to accomplish all this, and direct the provisioning of the forces—in essence giving complete marching orders to the military. At the other extreme, civilians could simply tell the military, “Deliver us from our enemies,” and let the military decide all the rest. The former would be tremendously costly—not to mention risky, if the civilians are incompetent. In the extreme, overmeddling could so jeopardize the lives of the military, or the fate of the mission, that the military would turn in revolt. Overdelegation would be the least burdensome and would avoid a *de jure* coup, but it would amount to a *de facto* coup: the military would be deciding policy and making decisions that by rights belong to the civilian political masters. This is precisely the fear of Kohn (1994) and Weigley (1993) with respect to the post–Cold War United States. Overdelegation requires the greatest trust in the military and leaves civilians vulnerable to the agency problems discussed above.

The military can be monitored, therefore, by restricting the scope of delegation.³⁰ At least insofar as military operations go, this type of monitoring takes the form of rules of engagement, standing orders, mission orders, and contingency plans. Rules of engagement, in principal-agent terms, are reporting requirements concerning the use of force. By restricting military au-

tonomy and proscribing certain behavior, rules of engagement require that the military inform civilian principals about battlefield operations whenever developments indicate (to battlefield commanders) that the rules need to be changed. Rules of engagement, then, are both a leash on the military and an information source for senior leaders, civilian and military. So long as the military operators do not “pull” on the leash, the senior commanders know that the pace of the military operation is less than the bounds set by the rules. The more restrictive the criteria, the more closely senior commanders can monitor the military operation (Sagan 1991). Similar dynamics obtain with standing orders, mission orders, and contingency plans, all of which are forms of detailed guidance on how operations are to be carried out (Bouchard 1991, pp. 32–34).

While all monitoring mechanisms can be more or less restrictively implemented, some are inherently less intrusive on military autonomy. The least intrusive forms of monitoring concern the designing of the contract that establishes the principal-agency relationship in the first place. Most economic applications stress controlling the agent through contracts that give agents an economic incentive to perform in ways that the principal wants; giving an agent a financial interest in the firm’s residual—that is, profit sharing—is a classic example. Political scientists have had difficulties applying this solution to political situations of agency because there is no obvious “profit” to be distributed.³¹ Profit sharing is straightforward when the organization’s basic output is easily measured (number of widgets) and the main goal of the principal is greater efficiency (more widgets at less cost). In political applications, the output (policy) itself can be in dispute or only imprecisely measured, and economic efficiency (more policy at less cost) is not the only desideratum. Political actors are sensitive to costs, however, and so a potential political analog is something called “slack,” the difference between the actual budget appropriation and the minimum cost of providing the service. Slack can be used to buy things that the agent (bureaucrat) wants, like new equipment, perquisites, and so on, but does not actually need to provide the service (Moe 1984, p. 763). In this way, the agent has an incentive to be efficient in providing the desired service, since he can spend the slack on things he values. Giving bureaus a fixed amount and allowing them to allocate it as desired is the most common use of slack in political organizations. The problem with using slack as a control mechanism is twofold. It requires that the principal consistently overpay for the service. And more important, slack does not solve the problem of ensuring that the policy output in fact

accords with the principal's desires and is not simply coming in at a lower cost.

The principal-agent perspective suggests that contractual incentives should be at the heart of the control relationship between civilians and the military. In the absence of an obvious profit to be shared, however, and given the limitations of slack as a surrogate, what are plausible incentives? Part of the problem of incentives is addressed through screening mechanisms (discussed below) that serve to populate the military with people who share, so much as possible, civilian preferences over outcomes. There are also historical examples of crass versions of profit sharing as a means of ensuring civilian control: the Romans, for instance, essentially bribed the capital garrison to keep it out of politics.³² Wages and benefits, and the implied sanction of withholding them, may be modern equivalents.

A particularly intriguing incentive, however, can be found in traditional organization theory's premise that organizations (agents) prize autonomy. Autonomy is slack without a monetary denomination. Since monitoring mechanisms vary in their degree of intrusiveness, and assuming that the military prefers less intrusive means, civilians have a powerful incentive with which to influence military behavior: offer to use less intrusive means to monitor military agents. Indeed, this is how traditional civil-military relations theory treats autonomy. It is the centerpiece of Huntington's ideal-type objective control and is even supported by fierce critics like S. E. Finer (Feaver 1993, 1996a; Clausewitz 1976, pp. 605–606; Brodie 1973, p. 494; Huntington 1957, pp. 83–85; Smith 1951, pp. 50–51; Finer 1962, pp. 39–56, especially 47–56, and 141–144; Betts 1991, p. 10; and Hendrickson 1988, p. 11). Claude Welch describes it as a “*noli-me-tangere*” approach: civilians promise autonomy to the military in matters of lesser import as an incentive for military acceptance of the ethic of subordination (Welch 1976, pp. 33 and 318).

A slightly more intrusive form of monitoring involves using screening and selection mechanisms to ensure that only the right sort of agent enters into the contractual relationship. This directly addresses the adverse selection problem, but it may be thought of as a relatively unintrusive information-gathering device. The way to make sure you have not hired a lout is to identify the characteristics of people who are *not* louts and then hire only them. Once you “know” the type of agent you have, you should be able to predict his behavior with greater confidence. Education requirements, skill tests, and problem-solving exercises all represent attempts by employers to screen

undesirables out of the candidate pool. Importantly, all traditional civil-military relations theories emphasize screening mechanisms of one sort or another, and the professional military education system is devoted to this kind of screening and socialization (Masland and Radway 1957).

Accession policy, determining who can join the military and how, is the military version of a screening and selection mechanism. Of course, the military uses elaborate physical, emotional, and mental tests to weed out poorly qualified applicants; this directly addresses the civil-military problematique by ensuring that the military will have the physical capacity to defend the country as needed. Accession policy can also mitigate civil-military problems by selecting or promoting personnel who will share civilian preferences. For instance, one of the primary arguments in favor of conscription is that it helps leaven the military mind-set with a steady supply of nonmilitary personnel, citizen-soldiers whose primary sense of identity and loyalty will be with civilian society. And by law in the United States, civilians have a fair amount of influence over the makeup of the officer corps. Congress votes on all officer promotions, and the more senior and influential the promotion, the more carefully the applicant is screened by Congress. Likewise, the president and his civilian staff often personally select the officers to be promoted to the most senior and sensitive posts, and in this way they shape the collective preferences of the officer corps. Changes of administration can result in changes in the degree of convergence between the officers appointed by a previous administration and the incoming civilian leaders (Whitworth and Watson 2001).

Compared with agency relationships in the other sectors of the bureaucracy, however, civilian principals have less discretion in using screening and selection to choose military agents. The president has virtually no limits on his selection of civilian foreign policy advisors, but he is limited to a finite number of senior military officers when he picks his most trusted military advisors. Of course, the president need not follow strict seniority precedence, and every president has exercised at least some latitude in promoting military advisors whom the president believes will be more in harmony with his administration's policies (Halperin 1972, p. 310). Janowitz found that the lower the rank of the military officers, the more frustrated they were with civilian viewpoints, and this suggests that promotion and selection procedures winnow out people who do not like civilian control, or at least that the views of those who get through the process are shaped by the organization (Janowitz 1971, p. 368; Schoenberg 1971).

The ability to shape viewpoints leads to yet another way that screening can serve to monitor delegation: through organizational culture. Organizational culture serves to provide mutual conceptions of behavior, allowing actors in a political game to have shared expectations of what the other will do.³³ Brehm and Gates found that these cultural factors, which they variously called “cohesion,” “solidarity,” and “professionalism,” had great explanatory power in determining when subordinates would work or shirk (Brehm and Gates 1992a, 1993). In the civil-military context, an organizational norm that stresses obedience gives both civilians and the military a common expectation that the military will be subordinate. In this one respect, the principal-agent framework shares a finding of traditional civil-military relations theory: that a common culture of subordination—Huntington’s “professionalism,” Welch’s “cult of obedience,” Smith’s “norm of civilian control,” Hendrickson’s “ethic”—is a crucial component civilian control (Huntington 1957, pp. 70–78; Welch 1976, p. 33; Smith 1951, p. 5; Hendrickson 1988, p. 26).

Beyond selection and screening mechanisms, the next most intrusive form of monitoring involves the use of third parties to watch the agent and report on key outputs, called fire alarms (McCubbins and Schwartz 1984). A third party who has a vested interest in the actions of the agent—for example, an interest group or affected constituents—can set off an “alarm” to alert the principal whenever the agent misbehaves. Thus the Sierra Club watches the Environmental Protection Agency, the American Association of Retired People watches the Department of Health and Human Services, and so on. The interest group environment is different in the military sphere, where there are fewer mass public organizations independent of the government, but there are many defense-oriented think tanks that investigate and report on the activity and adequacy of the defense establishment. It is also possible to think of conscription-based accession policies as playing something of a fire-alarm function: draftees, as resident civilians whose primary identity and loyalty is with civilian society, may be expected to sound the alarm if things are going awry.³⁴

The most prominent fire alarm on defense policy is the news media. The media act as independent third agents, self-appointed public watchdogs. Anecdotal evidence suggests that the media play an important role even for senior policymakers who have a wealth of internal sources of information. Reading the *Early Bird* clipping service, a daily compilation of all Department of Defense-related articles in the major newspapers, is usually a top

priority for senior policymakers; within the Pentagon, the joke is, “If it isn’t in the Early Bird, it didn’t happen.”

Interservice rivalry operates as another, slightly more intrusive fire alarm. Traditional civil-military relations theory has long recognized that the separateness of the American military services can bolster civilian control. In monitoring military behavior, a sister service may not face the same information asymmetries as would a civilian overseer. Civilian intervention can be triggered when one service complains about another either to the secretary of defense or to Congress. Civilian principals, lacking at least some relevant military expertise, cannot be sure that any particular military advice is sound. Having separate military services available to “sound off” and provide alternative military opinions is a guard against this.³⁵ This function was explicitly recognized in the congressional debates around the military unification movement immediately after World War II. Defenders of the status quo (strong separate services) touted the importance of alternative military voices. While the National Security Act of 1947 did centralize the military establishment in one Department of Defense, it preserved at least the framework of ongoing service rivalry in the separate service departments, each headed by a civilian service secretary, so as to make sure that alternative voices were heard.³⁶

Interservice rivalry can also be thought of as an institutional check. Institutional checks are related to fire alarms, but the principal-agent literature usually treats them as distinct. An institutional check is a separate agent, established by the principal and empowered with a veto to block action of the other agent. The function of a simple fire alarm is to alert the principal, who will then intervene to punish or adjust behavior as needed. The function of an institutional check is more assertive—to block, either legally or in some cases physically, any behavior that might be considered untoward. Institutional checks play a key role in civil-military relations. They are integral to what Huntington calls “subjective control,” and they have proven important in preserving military subordination to political authorities in ethnically divided states (Huntington 1957, p. 82; Frazer 1994; Horowitz 1985; Belkin 1998). In the U.S. case, the classic institutional check for the use of force is the Constitution’s division of military decisionmaking authority between the executive and the legislative branch; here, the framers of the Constitution treated the electorate as the principal and government as the agent. By extension, the presence of senior civilian officials in the Department of Defense, over which Congress, through the Senate confirmation process, has

some control, is also an institutional check. Likewise, the civilian staffs of congressional committees are another important institutional check, since they can block the activity of the executive branch. The framers clearly intended institutional checks to be the bulwark of civilian control over the armed forces; the purpose of a separate militia (now the National Guard) was to be a last line of defense against the regular standing army, should it prove bent on a usurpation of control.

The role of the militia and the National Guard has declined more than the framers would have expected, however, and to a certain extent interservice rivalry can be thought of as a replacement check. To the extent that the existence of separate services makes carrying out a coup that much more difficult, the services can be treated as separate sub-veto groups. At least in the U.S. case, however, interservice rivalry has played a more important role as an information gathering device for civilian principals, so I treat it as a traditional fire alarm rather than as an institutional check. The fire alarm function of interservice rivalry feeds back into institutional checks in an interesting way. Congress has historically been the most interested in preserving interservice rivalry, because the information asymmetries hit Congress especially hard. Congress wants to make sure that independent service chiefs will be able to come to Capitol Hill and disagree with each other, and especially with the administration, when necessary (Scroggs 1996).

Congress's institutional interest in interservice rivalry makes the push for the 1986 Goldwater-Nichols' reforms especially interesting from a principal-agent perspective (Weiner 1997). One of the main goals of Goldwater-Nichols was to weaken interservice rivalry by strengthening the ability of the Joint Chiefs of Staff (JCS) to provide a joint military viewpoint. In so doing, Congress was putting a muzzle on the interservice rivalry fire alarm. There are several reasons why it may have done this. First, Congress may not have thought the reforms would entirely silence interservice disagreements. Indeed, the reforms explicitly preserved the service end-run option to Congress, so the muzzle was not so restrictive as it may at first glance seem. Second, interservice rivalry had proven a less-than-ideal fire alarm from Congress's viewpoint. The service chiefs had learned to logroll among themselves without making real strategic trade-offs. Then they would present a united front to Congress and the president, and this would force civilians either to spend too much on the military function or make ill-informed cuts (Hoopes 1954, p. 228). Third, perhaps Congress viewed the newly strengthened Joint Staff as creating yet another fire alarm, rather than

dampening existing ones. The Joint Staff would watch the services closely and could provide a strong warning if the services colluded or shirked in untoward ways.

The foregoing suggests that principals are sensitive to the costs of fire alarms, and these costs extend beyond the intrinsic costs emphasized in the literature: the likelihood that the fire alarm will lie or will be unable to accurately report on what the agent is doing (Lupia and McCubbins 1994). The civil-military story points also to extrinsic costs. For instance, interservice rivalry may provide useful and valid information to civilian principals but only at a cost, such as potentially needless duplication and interoperability problems. Likewise, the media can be a useful fire alarm when it reports on peacetime abuses involving hazing or sexual harassment. But in wartime, the media may not be the kind of fire alarm on which civilians will want to rely because, in keeping with the metaphor, both firefighters and arsonists can learn from a fire alarm; if the media report that the military has not adequately defended a certain sector, this is useful information for the principal but it is also useful information for the enemy. Hence the ubiquity of media restrictions during military operations (Smith 1992, Bennett and Paletz 1994).

Public fire alarms like the news media also have another cost deriving from the fact that the civilian policymaking principal is himself an agent of a still more ultimate principal, the voter. A fire alarm that alerts the policymaking principal that something is awry also alerts the voter that something has gone wrong on the policymaker's watch. For this reason, the policymaker has an incentive to prefer a monitoring mechanism that produces reliable information *privately*, or at least in a way that can be shielded from the voter.³⁷

Institutional checks also have associated costs. Institutional checks work best when the interests of two agents are in conflict, either because they face different contract incentives (one is paid for cutting costs, the other for boosting production); otherwise, the two agents could collude and the principal would be back facing the moral hazard problem. Moreover, for an institutional check to be effective, each must have something akin to the power of a veto over the other. For these reasons, interservice rivalry is not a formidable institutional check. The services all are tasked with the same thing—providing security—and the services do not really have a veto over each other, except at the most rudimentary and theoretical level of being able to counterbalance each other in any military takeover attempt. Institu-

tional checks can be effective, but there are high costs associated therewith: they make it harder for the agent to do bad things, but they also make it harder for the agent to do good things.³⁸ Civilian principals have from time to time created new institutional checks in the evolution of American civil-military relations, but only when the stakes are particularly great. The most obvious example was the establishment of the Atomic Energy Commission to assist civilian control over nuclear weapons (Feaver 1992).

The next most intrusive form of monitoring has been dubbed “police patrol” monitoring (McCubbins and Schwartz 1984). This involves regular investigations of the agent by the principal—fishing expeditions, if you will, where the quarry is general information on what the agent is doing. Police patrols include regularized audits and intrusive reporting requirements designed to turn up evidence of agent wrongdoing and, through regularized inspection, to deter moral hazard. Public investigative hearings and specific mandated reports are staples of congressional oversight and represent one of the more visible avenues of political control. Similar mechanisms operate within the executive branch to facilitate hierarchical control and some, like the Planning, Programming, and Budgeting System in the Department of Defense, directly concern security issues. Congressional investigations have at times featured prominently in the history of American civil-military relations, particularly the Committee on the Conduct of the War during the Civil War and the Senate Permanent Investigations Subcommittee under Senator Joseph McCarthy. Likewise, reports and audits are ubiquitous in the politics of the defense budget.

In the civil-military context, an important indicator of police patrol monitoring is the size of the civilian secretariat of the Office of the Secretary of Defense and the service secretariats. These are extensions of the executive branch principals, the patrol officers, who are in place to monitor closely and directly the activities of their military counterparts. Accordingly, large numbers of civilian officials are evidence of a police patrol monitoring mechanism.

Inspectors general are a hybrid of police patrol and fire alarm monitoring. On the one hand, they are internal to the organization and have full audit authority; on the other hand, an inspector general’s investigation is not a regularized audit but is usually triggered by some precipitating factor, like a leak.

Other examples of police patrol monitoring in the civil-military sphere are the activities of the various governmental investigative agencies, the Con-

gressional Budget Office, the General Accounting Office, and the Office of Technical Assessment. While they bolster the strength of Congress vis-à-vis the executive branch, and so may be thought of as part of an institutional check, they function to increase access to military information by the civilian principals as a whole. The annual reports to Congress are likewise examples of police patrol monitoring; while their nominal audience is Congress, the reports are first reviewed by civilians in the executive branch, and the act of writing and reviewing them produces valuable information for monitoring the activity of the military establishment. In a different context, communications links from national command authorities to operational commanders constitute tangible monitoring and auditing channels. Also, as discussed above, restrictive rules of engagement that narrow the scope of delegated authority are examples of police patrol monitoring. Like traditional principal-agent oversight mechanisms, these measures are costly in terms of civilian attention (not to mention dollars) but can mitigate somewhat the informational asymmetries in the civil-military relationship.

At the most intrusive end of the oversight spectrum would be a decision to revisit the original decision to delegate authority to the military agent in the first place. Civilian principals have the option of redrawing the boundary, crossing over into the military zone to make or implement a decision on a particular issue or set of issues. For instance, President Johnson delegated responsibility to General Westmoreland to conduct the Vietnam War, and civilians oversaw that delegation through a prescribed set of reporting requirements, rules of engagement, and so on. Despite this arrangement, from time to time the president and his senior civilian advisors intervened in the war not only to tinker with the monitoring system but also to make operational decisions themselves. Interventions such as the celebrated practice of selecting bombing targets from the basement of the White House had the effect of directly narrowing the freedom of action of the military agent and so constitute an especially intrusive form of oversight.

Table 3.1 summarizes the various oversight mechanisms just discussed.

Of course, civilians would not rely on just one or two mechanisms but rather would use a mix. Moreover, the mix would be in most instances cumulative. Thus, if civilians sought to monitor the military more intrusively, they might bolster the civilian complement within the Office of the Secretary of Defense (add police patrolmen) without shutting down defense-oriented think tanks (preserving existing fire alarms). Similarly, a move toward less intrusive monitoring would be indicated not so much by the establish-

Table 3.1 Summary of oversight mechanisms in ascending order of intrusiveness

Monitoring mechanism from principal-agent literature	Civil-military analog
Contract incentives	Offer by civilians to use less intrusive monitoring in exchange for obedience
Screening and selection	Skill requirements for entrance into military Loyalty oaths Other accession instruments Professionalism
Fire alarms	The news media Defense-oriented think tanks Interservice rivalry
Institutional checks	Militia system and National Guard Interservice rivalry (sometimes) Civilian staffs in Congress Atomic Energy Commission Confirmable civilian secretariat
Police patrols	Planning, Programming, and Budgeting System and the budget process Civilian secretariat and Office of Secretary of Defense Restrictive rules of engagement Restrictive standing or mission orders Limits on delegated authority Audits and investigations Inspectors General Congressional Budget Office, General Accounting Office, Office of Technical Assessment
Revising delegation decision	Intervening in a military operation to make a decision that was hitherto in the scope of delegated authority (e.g., picking bombing targets from the White House)

Copyright © 2003. Harvard University Press. All rights reserved.

ment of new fire alarms as by a reduction in the more intrusive forms of monitoring. *Ceteris paribus*, the addition of a fire alarm without any other compensating action, would increase the degree of intrusiveness by adding sets of eyes to the monitoring mission.

Civilian Punishment of the Military

While attention to monitoring mechanisms is a hallmark of the principal-agent literature, punishment mechanisms are often only implicit. It is assumed that the only problem is detecting whether shirking is going on; once detected, punishment is automatic, or at least unproblematic. To the extent that the literature does consider postmonitoring issues, it is usually in the context of the reward incentives built into the original contract. The agent works in order to receive the rewards, and if the agent shirks, and this is detected, the withholding of rewards constitutes a form of punishment. Proactive punishments, materially reducing the agent's utility beyond the withholding of a reward may be only summarily mentioned (Kiewet and McCubbins 1991, p. 29; Lupia and McCubbins 1994, pp. 104–105; Milgrom and Roberts 1988, p. 157; McNollgast 1987, p. 252; and McNollgast 1989, p. 439).

A few treatments make punishment more explicit and have it play a more central role in ensuring compliance. For instance, Bianco and Bates's analysis of iterated games found that the leader's ability to set punishments was critical to the initiation and sustainment of cooperation (Bianco and Bates 1990, Pollack 1997). Similarly, principal-agent games between an electorate and a leader turn on the ability of the electorate to vote out of office, or punish even more severely, leaders who fail them (Goemans 2000, Downs and Locke 1995, Bueno de Mesquita and Siverson 1995). Lindsay compared congressional oversight of domestic and foreign policy and concluded that the difficulty Congress had in punishing foreign policy actors was a primary factor in explaining the diminished congressional control in that arena (Lindsay 1994a, pp. 283 and 298).

One important application of the principal-agent framework, however, has explicitly challenged the role of punishment. John Brehm and Scott Gates claim that credible threats of punishment from supervisors are inefficient means of gaining compliance.³⁹ They argue that compliance is far more efficiently gained when agents share the preferences of the principal. Principal-agent models have generally assumed widely divergent prefer-

ences—principals want agents to work, and agents want only to shirk—and Brehm and Gates dismiss this as an unproven assumption “cloaked in theological robes” (Brehm and Gates 1997, p. 20). I accept part of their critique; as explained above, in civil-military applications of the principal-agent framework one must incorporate a richer spectrum of military agent preferences, including the possibility that the preference gap will be narrow under certain circumstances. But to show that agency problems are diminished when principals and agents share preferences is to show that people who agree with each other tend to cooperate. Downs and Rocke strongly criticize this view and argue that the cooperation one gets in this way tends to be shallow (Downs and Rocke 1995). *Some* divergence of preferences some of the time is inevitable and so the moral hazard problem is not completely avoidable.⁴⁰ If there is moral hazard, there is at least the potential for punishment, a consequence of particular interest in the civil-military context.

Punishment has received an uneven treatment in the civil-military relations literature as well. On the one hand, the military has been distinctive as an organization with a high reliance on coercion to enforce discipline—as in the famed Churchillian trinity of the British Navy: “rum, sodomy, and the lash.” Likewise, harsh punishments for battle cowardice *pour encourager les autres* was long an accepted means of forging an effective fighting force. Thus the role of coercion, at least insofar as it concerned obedience *within* the military organization, has always been a part of the military organization (Bryant 1979, pp. 34–35). Military sociology has for years debated whether rigid enforcement of rules, often with extreme physical coercion, is more effective in enforcing unit obedience than softer forms of manipulation, persuasion, and group consensus (Janowitz 1971, pp. 38–53; Henderson 1985, pp. 9–26).

On the other hand, punishment is rarely discussed in the theoretical literature on civil-military relations as a relevant tool in enforcing discipline *across* the civil-military divide. Punishment is not emphasized in the theoretical literature partly because the military uniquely controls coercive power. If the civilian tries to punish the military, what is to stop the military from resisting by force? One of the distinctives of the civil-military relationship is the fact that the subordinate is almost always more powerful than the superior. This is always true in the most basic sense of brute force. It can even be true for more intangible measures of power. The military may have tremendous political power because it is an important consumer block in a market economy. Likewise, the military can enjoy a prestige that confers political power quite apart from any consideration of physical coercion. Given the

overall power differential, an exclusive focus on punishment would seem misguided. Yet, as I will argue below, civilians *can* punish the military, and sometimes they do.

The relative inattention to punishment may also be explained by the orientation of the theoretical literature on which most of the civil-military relations literature draws. Both the Huntingtonian and Janowitzian schools essentially accept the neoclassical critique of classical organization theory.⁴¹ The classical organization theory of Frederick Taylor and Max Weber saw coercion and material incentives as an essential element of any bureaucracy. The neoclassical work of Carl Friedrich, Chester Barnard, and Herbert Simon was a reaction against the material and coercive basis of organizational behavior and emphasized instead the role of professionalism, cooperation, and bounded rationality. Huntington adopted the neoclassical emphasis on professional norms as the determinant of behavior, and Janowitz likewise embraced socialization to codes of behavior as the key to military obedience.⁴² The empirical foundation of the neoclassical rejection of coercion and punishment is not unassailable, however. A limitation of all the studies that show coercive power is relatively ineffective is that they measure compliance based on self-reports.⁴³ The most accurate summary of existing empirical data, then, would be: respondents who perceive that their superiors are relying on professional norms and persuasion rather than coercion to induce compliance are more likely to report that they comply with their superiors' orders.⁴⁴ This is an important finding, but it makes something less than a compelling case that coercion either plays no role or ought to play no role.

Therefore, I explicitly consider the role punishment plays in U.S. civil-military relations. In so doing, however, I sidestep a problem that arises in some civil-military settings but not in most instances of principal-agent relations: whether civilian principals *can* punish the military agents. There are numerous examples of coups triggered by a civilian decision to punish some part of the military for an earlier disobedience. Such a coup effectively neutralizes any ability to punish. At some level this problem is simply assumed away by the principal-agent framework. As discussed above, agency theory is only applicable in those settings where the military conceives of itself as the agent of the civilian; crucial to that conception is a recognition of the civilian's right to sanction, and hence an explicit commitment to submit to sanctions. Such an assumption is reasonable in the U.S. case. There is ample evidence that American civilians are able to punish the military, if they so choose. Civilians have the ability to fire even hugely popular military officers, as Truman's dismissal of MacArthur makes clear. Many senior military officers

have been sacked before and since for a wide range of offenses that can be grouped collectively under the heading of shirking. But it bears emphasis that the power to punish rests on a normative foundation—that is, the willingness of the military to be punished—and this normative foundation is thus a prerequisite for democratic civil-military relations. It exists in the United States and other advanced democracies but not necessarily in all countries. As discussed in the concluding chapter, this may limit the applicability of the agency model to other countries.

In democracies, therefore, civilians *can* punish the military. Nevertheless, *whether* civilians will use their ability to punish is uncertain—MacArthur evidently did not think it was guaranteed that Truman would punish him. The principal-agent literature acknowledges that punishment is a costly action by the principal and is by no means assured in every instance of agent shirking. The primary finding from the literature is that when there are multiple principals, a shirking agent can play one off the other and thus reduce the likelihood of being punished (McNollgast 1989, p. 439). In the U.S. civil-military context, the ability of the military to play Congress against the president and vice versa is an obvious analog (Avant 1994, p. 14).

Thus, in addition to the uncertainty over whether the behavior will be discovered, there is uncertainty over how the alleged shirking behavior will be interpreted by the civilian principal. Most principal-agent applications assume that the behavior is hidden or unobservable, but that the character of the behavior would, if known, be unambiguous (Downs and Rocke 1995). In most instances, this is a reasonable assumption. In the civil-military context, however, it is plausible that the nature of the behavior itself is ambiguous, subject to different interpretations by different civilian principals. What is excessive force in combat? How much candor can senior military display in their testimony before Congress when they disagree with administration policy? Some activities might be obvious—collecting war booty, for example—but the norms governing the acceptability of the behavior may change (Schmitt 1992a). Civilian principals have the right and the ability to set the boundary of appropriate behavior and to interpret ambiguous behavior as they see fit. Obviously, information about what really happened, as well as information about any extenuating circumstances, will partly influence the perception of the behavior as shirking or not. But it will also depend on other factors, including calculations of costs not unlike those that I argue govern monitoring and shirking decisions: the salience of the issue, the popularity of the offending military agent, and so on.

If civilians decide to punish the military, they can select from an almost

infinitely wide range of punishment tools, which may be grouped into five broad categories. The first involves imposing the kind of monitoring arrangements the military dislikes. As noted above, the military has a strong preference for nonintrusive monitoring. If in response to some outbreak of shirking the civilians impose an intrusive monitoring regime, it constitutes something of a punishment, albeit not a draconian one. Of course, it is difficult analytically to distinguish this kind of punishment from the prior decision to monitor. Such monitoring is often experienced as punishment by the military agent—consider, for instance, the heightened public scrutiny of the way the military has integrated women since the U.S. Navy’s Tailhook crisis—but its intended function may fit better under the heading of monitoring than sanctioning.

The second set of punishment mechanisms involves cutting budgets and reducing the perquisites enjoyed by the military. In the civil-military context, this is most often seen as a by-product of a standoff between the congressional and executive branches. In such cases, Congress may make compensating cuts in appropriations or temporarily hold up promotions in order to force some compliance with a policy that the executive branch is resisting. I do not know of instances when this was used as a punishment of the military by the civilian executive branch, but it would seem plausible.

The third set of punishments, and by far the most prevalent, involves variations on forced detachment from the military—the military equivalent of firing. The military is distinctive as a profession with both an up-or-out career path and a very generous prize for those who stay in a “full” term, historically twenty years. The rewards for “making it to retirement” are substantial—a guaranteed pension equivalent to roughly 50 percent of one’s last base pay, payable immediately—and they are won after just twenty years of service. Moreover, the rewards climb exponentially with seniority. The longer one serves, the greater the preretirement baseline salary and also the greater the percentage of that baseline salary one receives as retirement pay. The fully vested retired military officer can be as young as forty-one (a fully vested enlisted person can be under forty), young enough to start a second career while collecting military retirement pay. And there is also a vast array of other amenities, like subsidized shopping services and subsidized medical care. There is more than a little irony in the fact that the military that prevailed against world communism is also the American institution that most closely approximates the idealized benefits of a socialist society. But these rewards accrue only to someone who makes it to retirement, that

is, someone who successfully earns each successive promotion and so is not discharged before twenty years are served. In theory, the up-or-out trigger is meant to ensure that underperformers do not last long enough to win the benefits. Retirement pay thus works as a carrot to entice the better soldiers and sailors to remain in the military rather than seek more immediately lucrative careers in civilian society, and also as a stick to enforce compliance, lest one is prematurely discharged from the service and thus denied a plum benefit.

The retirement benefit can be manipulated as punishment in several ways. Most obviously, a deviant officer can be discharged from the military for cause. If the officer has not yet served twenty years, he or she loses the most important service benefit, retirement pay. A slight variation on this involves failing to promote an officer. Officers serve with a time-limited commission. They become eligible for promotion by accession-year cohorts, according to a strict and linear time-in-service calculation. A promotion board of more senior officers, using guidance approved by the senior civilian leadership in each of the services, decides which officers in a cohort will be promoted and which will not. Generally, candidates not selected for promotion the first time they are eligible have one more chance at the promotion. "Twice passed-over" candidates must leave the service when they are ordered to do so. During the Cold War, a steady influx of junior officers pushed out those who had not been promoted to higher ranks. Not promoting an officer can thus be equivalent to firing that officer.⁴⁵ Competition for promotion intensifies the more senior the officers get; consequently, the more sensitive those officers become to having adverse information in their personnel record.

Another variation involves the ability to retire an officer at a lower rank than the highest one he achieved. This can be done directly by the service secretary, as an explicit punishment for behavior that otherwise did not rise to the level requiring a formal discharge. It can also be done indirectly via a rule that requires officers to serve a certain length of time at a rank before that rank can be the basis for retirement pay calculations. If they retire before they have served long enough to claim that rank for retirement purposes, in effect they retire below grade. This is of particular significance for the highest ranking officers, admirals in the navy and generals in the other three services, because these ranks are tied to specific billets and assignment to those billets may be for a shorter term than is required to guarantee the retirement privileges of the rank associated with that billet. If they lose the

billet (for example, commander of the Seventh Fleet) before the requisite two years and are not assigned to another billet requiring the same rank, they will have to retire at a lower grade, with correspondingly lower pay.

The final variation is the more well known system of discharges. There are three administrative discharges: the honorable discharge (under which all privileges are maintained), the general discharge (an intermediate discharge reflecting some minor disciplinary infraction, under which certain benefits may be maintained), and the general discharge under other than honorable conditions (a more severe dismissal, usually given in lieu of a court-martial, under which most benefits are lost). And there are two punitive discharges that carry the stigma of a felony conviction, that can be issued only by a court-martial, and that deny virtually all benefits: a bad conduct discharge (which can be issued by a special court-martial) and, the most severe of all, a dishonorable discharge (which requires a general court-martial).

The fourth set of punishments involves the complex system of military justice as specified under the Uniform Code of Military Justice (UCMJ). This can involve the separation of the service member from the military, but it need not. The UCMJ has provisions for capital punishment, imprisonment, or simply the military equivalent of fines and community service. The UCMJ and its historical antecedent, the Articles of War, have not primarily been instruments of civilian control, although the military judicial system has relevance for military subordination. The UCMJ has first and foremost functioned as an instrument for maintaining command discipline *within* the military—that is, as a tool for senior military commanders to use in controlling the behavior of their military subordinates (Lurie 1992). Nevertheless, to the extent that civilian leaders determine which behaviors are proscribed by law and which areas are left to commanders' discretion, and to the extent that senior military officers take their cues about exercising this discretion from civilian leaders, the military justice system can be considered a part of the civilian monitoring and punishment edifice.

The fifth set of punishments involves extralegal civilian action taken against specific military personnel. This is a miscellaneous category of actions ranging in severity from private oral rebukes all the way to the infamous Stalinist purges against the Soviet military in the 1930s in which thousands of officers were shot for suspected disloyalty to the Soviet regime. An intermediate form might be a situation in which the military advisor is publicly reprimanded or denied access to the civilian leader because that leader has lost confidence in him. What distinguishes this as a means of punish-

ment is that it is a direct expression of civilian supremacy and is usually directly tied to the kinds of shirking behavior of interest to civilian principals.

One might argue that the domain of civil-military relations offers yet a sixth form of punishment that does not appear in domestic principal-agent settings: war. War, of course, inflicts hardships on the military, including the possibility of the ultimate sanction, loss of life, and one might consider war as a form of punishment. If war performance is itself a function of civil-military relations, then battlefield defeat could be a form of punishment experienced by the military (and, less directly, by the civilians as well) for adopting suboptimal civil-military arrangements. Intriguingly, Cohen (2001) argues that, contrary to a cherished military view, war performance improves with vigorous civilian involvement in the details of the war (what the agency theory considers to be intrusive civilian monitoring). In this sense, military shirking or resistance to intrusive monitoring might increase the risk of battlefield collapse and thus constitute a punishment of sorts. War is rare enough that it probably does not form a primary punishment vehicle anyway, and it certainly does not preclude a role for the more mundane forms of punishment discussed above and summarized in Table 3.2.

Table 3.2 Military punishments available to the civilian principal

General category	Examples
Restrictive monitoring	Audits Mandatory remedial training (as in sexual harassment training)
Material disincentives: current	Budget cuts Restricting liberty
Material disincentives: future	Discharge prior to earning benefit Retirement below grade Loss of retirement privileges Other-than-honorable discharges
Military justice system	Nonjudicial punishment Courts-martial
Extralegal action	Verbal rebukes Purges

Copyright © 2003, Harvard University Press. All rights reserved.

Conclusion

The civil-military relationship is at its heart an agency relationship, and so the principal-agent framework developed in microeconomics and already used in various political applications can be profitably extended to the study of civilian control of the military. The civilian principal establishes a military agent to provide the security function for the state, but then must take pains to ensure that the military agent continues to do the civilian's bidding. Given the adverse selection and moral hazard problems endemic in any agency relationship, but particularly acute in the civil-military context, civilian oversight of the military is crucial. Fortunately, civilians have available a wide variety of oversight mechanisms, each involving a different degree of intrusiveness and therefore each posing a different set of costs on the actors. The oversight regime is supported by a sanction regime, which provides civilians with options for punishing the military when it shirks, that is, deviates from the course of action prescribed by civilians. This basic story is analyzed in a formal game in the next chapter.