

DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES

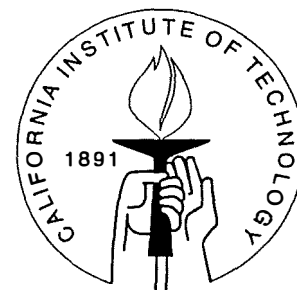
CALIFORNIA INSTITUTE OF TECHNOLOGY

PASADENA, CALIFORNIA 91125

DELEGATION AND THE REGULATION OF RISK

Hugo Hopenhayn
University of Rochester

Susanne Lohmann



SOCIAL SCIENCE WORKING PAPER 993

August 1992
Revised July 1996

Delegation and the Regulation of Risk

Hugo Hopenhayn Susanne Lohmann

Abstract

Political principals typically use low-cost “fire-alarm” signals transmitted by the media, interest groups, and disaffected constituents to monitor the activities of regulatory agencies. We argue that regulatory decision-making is biased and inconsistent if the instruments of political oversight are simple and the information flows to the principal are coarse relative to the complexity of the regulatory environment. Journal of Economic Literature Classification Numbers: D72, L51

Delegation and the Regulation of Risk*

Hugo Hopenhayn Susanne Lohmann

1 Introduction

U.S. regulatory agencies often regulate or ban economic activities that are associated with minor risks, while delaying action or not enforcing regulations for more hazardous activities. Environmental health and safety standards change over time and differ across agencies or even across programs controlled by the same agency in ways that appear unrelated to the underlying scientific “fundamentals.” In many cases, the pattern of regulatory response is one of overreaction to sensational accidents and disasters (Magat et al., 1986; Viscusi, 1992; Office of Management and Budget, 1992; Wildavsky, 1995).

The tendency of regulatory agencies to set inconsistent standards and to overreact to random incidents is puzzling. Efficiency considerations suggest that more injuries and premature deaths could be prevented if the stringency of regulatory standards were equalized across hazardous activities. Occasional random accidents or disasters should be accepted as an inevitable by-product of economic trade-offs in the regulation of risk.

One explanation for this puzzle is founded on bounded rationality (Slovic et al., 1985). Both casual introspection and a large body of experimental evidence suggest that people are not fully rational Bayesians. Their inferences are subject to cognitive biases that arise from the use of suboptimal risk assessment heuristics; and they use heuristic rules-of-thumb in processing new information (Kahnemann et al., 1982).

* We thank Kathleen Bawn, Jonathan Bendor, David Baron, Linda Cohen, Lester Lave, Mathew McCubbins, and Eric Rasmusen for useful comments. Earlier drafts of this paper were presented at the First Conference of the Society for Public Choice and Welfare in 1992, the Meetings of the Midwestern Political Science Association in 1993, Stanford University, and the University of Indiana, Bloomington.

In particular, people tend to overestimate the frequency of low probability events and underestimate the frequency of high probability events. They also have little appreciation for the law of large numbers and tend to overgeneralize from small, potentially unrepresentative, samples. One implication would be the observed pattern of alarmist reactions toward infrequently realized hazards, on the one hand, and relative indifference toward frequently realized hazards on the other. Even if public policy experts are aware of the nature of cognitive processes and can “undo” biases in their own risk assessments, they might be forced to go along with the wishes of the public and its untrained intuitions.

Implicit in this view is the notion that citizens’ preferences are more or less directly translated into public policy. In fact, the preferences of the public are mediated through political processes and institutions. It is clearly of interest to examine whether political factors contribute independently to the puzzling empirical regularities noted above.

This paper deals with the question whether the over- and underregulation of hazardous activities may be due to some characteristics of the political oversight process. The activities of regulatory agencies are typically not monitored in an ongoing fashion. Instead their political principals--Congress, Congressional committees and subcommittees, the White House--rely on external information flows about the risks associated with various regulated activities. These principals are regularly informed about dramatic, well-publicized accidents and disasters as well as complaints filed by disaffected interest groups and constituents. Congress may respond to such “fire-alarm” signals by engaging in oversight activities (McCubbins and Schwartz, 1984). It has access to a variety of budgetary and legislative instruments of political control. Perhaps most importantly, Congress can override expert judgments or decisions made by agencies, for example by initiating new legislation or withdrawing funding for implementation. In the case of executive branch agencies, the president may have the option of dismissing agency appointees, or “encouraging” them to resign.

So far we have described one level of the political oversight process: the principal-agent relationship between policymakers and regulators. Another level is given by the principal-agent relationship between voters and their elected representatives. These two levels of oversight resemble each other in significant ways. The primary instrument of oversight employed by voters is dismissal: if they are dissatisfied with their political

representation, they can vote the incumbent out of office. In some cases, voters have access to another instrument of oversight: the referendum, which can be used to overturn political decisions. Voters also have access to fire-alarm signals--specifically media reports, interest group communications and, in the case of elections, public statements made by challengers (Lupia, 1992, 1994).

A more complex analysis would explicitly model this situation as a two-level, or nested, principal-agent problem. For simplicity, we focus on the principal-agent relationship between a policymaker and a regulator, referring to the former as the principal, to the latter as the agent. We examine the case in which the principal can override the agent's decision or dismiss the agent in response to fire-alarm signals. It should be understood, however, that our analysis also applies to the principal-agent relationship between voters and policymakers.

Two characteristics of the political oversight process are central to our analysis. First, when delegation is motivated by differences in expertise or costs of information gathering, the principal tends to be less well informed than the agent. Even if the principal has access to low-cost fire-alarm signals, the information conveyed by the media, interest groups, and constituents is often very simple in nature. As a consequence, the principal tends to have a coarser information set than the agent. We formalize this notion by assuming that the agent observes a continuous variable (the probability that a hazard will occur if a hazardous activity is allowed), while the principal only observes two binary pieces of information (the agent's decision to allow or prohibit the hazardous activity and a favorable or adverse fire-alarm signal).

Second, political oversight instruments are simple relative to the complexity of the regulatory environment. The principal's choice set is discrete: she can make a binary choice between overturning the decision of an agent and not doing so; or dismissing the agent and not doing so.

The principal's discrete choice set and the coarseness of her information set are two important sources of regulatory bias. If the principal had access to a richer set of signals or oversight instruments, she could fine-tune her political control rule to ensure that the agent's behavior is more in line with her objectives. Our analysis suggests that we should expect regulatory inconsistency and overreaction to be pervasive, albeit without relying on

bounded rationality considerations. More specifically, low probability hazards will tend to be overregulated, high probability hazards underregulated.

Other scholars have identified two features of fire-alarm oversight that may prevent information pertinent to political decisions from being fully aggregated. First, if there are multiple information providers and the political alternatives under consideration have the characteristics of a collective good, then the provision of costly informative signals is subject to a free rider problem (Lohmann, 1993). Second, information providers whose interests conflict with those of political decision-makers may have incentives to deceive the recipients of their signals (Lohmann, 1993; Lupia, 1992). Our paper identifies a third feature of fire-alarm oversight that may prevent the “full-information outcome” from being achieved: fire-alarm signals and oversight instruments are simple relative to the complexity of the political environment. This third feature is of particular importance because it addresses the logic of fire-alarm oversight at its very core. Fundamentally, fire-alarm oversight makes sense in a complex environment where the acquisition of “encyclopedic” information is prohibitively costly but low-cost “information shortcuts” are possible (Lupia, 1994, p. 63).

This paper studies the regulatory implications of the political oversight process by analyzing the equilibrium of a regulation game. Rather than study a general formulation, we have chosen to keep this game as simple as possible in order to make transparent the intuition underlying our results. To this end, we abstract from many otherwise important features of regulatory decision-making and political oversight. While our model lacks richness, its abstract nature allows us to isolate the role of simple oversight instruments and coarse information flows in creating regulatory bias.

Section II develops a model of delegation in which the principal may override the agent’s regulatory decision. Section III modifies the model to analyze the situation in which the principal may dismiss the agent. Section IV discusses various applications and extensions of our analysis.

2 Delegation with Override

A society faces the regulatory decision to allow or prohibit a hazardous activity.¹ The activity generates some benefits for society. With some positive probability, an adverse state of the world is realized. If the activity is allowed and the state of the world is adverse, some members of the society suffer costly health consequences. Otherwise, no costs are incurred.

One example for the kind of regulatory decision we have in mind is whether to approve a drug that promises to fight disease but may be associated with negative side-effects. Another example is given by the regulatory decision whether to allow a nuclear power plant to be built in an area that may be prone to earthquakes.

We first develop a two-period principal-agent model, in which the political principal of a regulatory agency may overturn agency decisions. The principal's utility in each period is given by

$$U^P = \begin{cases} b & \text{if the activity is allowed and the state of the world is favorable} \\ b - c & \text{if the activity is allowed and the state of the world is adverse} \\ 0 & \text{if the activity is prohibited} \end{cases}, \quad (1)$$

where b are the benefits derived from allowing the members of the society to engage in the hazardous activity, $0 \leq b \leq 1$, and c are the costs then incurred in the event of an adverse environment. (Time subscripts are suppressed whenever possible to avoid notational clutter.)

The ex ante probability that the adverse state of the world will be realized is given by π , while the favorable state of the world is realized with probability $(1-\pi)$. The probability π is drawn from a uniform distribution on the interval from zero to 2Π ; that is, the mean probability is equal to Π . For the time being, we set Π equal to one half; but later on, a low

probability hazard will be associated with $\Pi < \frac{1}{2}$, and a high probability hazard with $\Pi > \frac{1}{2}$.

The principal is imperfectly informed about π .² Furthermore, it is prohibitively costly for the principal to acquire information about this probability. But she can costlessly provide a regulatory agency with the apparatus and the expertise to determine perfectly and costlessly the value of π . Since π is privately observed by the agency, the principal lets the agency decide whether to allow or prohibit the activity.³

Thus, we motivate delegation with the standard informational argument (Kiewiet and McCubbins, 1991). In practice, it is infeasible (prohibitively costly) for a political principal to acquire detailed information about the incidence and severity of the hazards associated with various activities. The regulatory agency has a comparative advantage in acquiring information. To the extent that the acquired information is private, perhaps due to its technical complexity, the agency would be granted decision-making powers. Our model can also be reinterpreted as pertaining to the case in which the agency makes an expert judgment whether to allow or prohibit the activity, and the principal makes the actual decision.

The principal thus delegates the authority to allow or prohibit the potentially hazardous activity to the regulatory agency, henceforth referred to as the agent. The agent is chosen from an infinite pool of candidates whose utility in each period is given by

$$U^\tau = \begin{cases} b + \tau & \text{if the activity is allowed and the state of the world is favorable} \\ b + \tau - c & \text{if the activity is allowed and the state of the world is adverse} \\ 0 & \text{if the activity is prohibited} \end{cases}, \quad (2)$$

where τ indexes the candidates' types.⁴

We specify a distribution of candidate types that is symmetric around the principal's type. The agent's type τ can take on one of three values, with equal probabilities: T , 0 , and $-T$. The parameter T reflects the heterogeneity of the pool of candidates and is restricted to the range $0 \leq T \leq \min\{b, 1-b\}$. This restriction implies that $b - T \geq 0$ and $b + T \leq 1$. We impose this restriction to preserve symmetry. If the distribution of preferences were asymmetric relative to the principal's preferences, our model would generate the implication that regulatory outcomes are biased. We are interested in analyzing other, less obvious, sources of bias and do not wish to confound the effects of various sources. In any case, we shall seek to make transparent our use of this restriction in the derivation of our results.

Without loss of generality, the costs c are normalized to one.⁵ In the absence of any other constraints, then, an agent whose benefits are given by $b + \tau$ allows the activity if the probability π lies below the cutpoint $b + \tau$, and prohibits the activity otherwise. The value $b + \tau$ will be referred to as the agent's myopic cutpoint. The cutpoint of an agent of type T is higher than the principal's cutpoint: the agent is willing to allow the activity for a larger range of π than the principal would. An agent of type 0 has the same cutpoint as the principal. Finally, an agent of type $-T$ has a lower cutpoint: he is more "cautious" than the principal. The parameter T thus captures the potential conflict of interest between the principal and her agent.

If the principal were completely informed about the preferences of the candidates in the pool, she would delegate power to an agent who is a clone of herself, that is, an agent of type 0 . In this case, the principal would never have incentives to interfere with her agent's decisions. The oversight problem becomes more interesting if the principal is incompletely informed about the candidates' preferences. She knows the distribution of their preferences, and in particular the parameter T , but she cannot directly observe the preferences of any specific candidate.

At first blush, it might appear implausible to assume that a political principal is incompletely informed about the policy preferences of a political appointee. It might be argued that the principal can form an estimate of a candidate's policy preferences on the

basis of personal contact, past decisions, interest group ratings, and the like. In our view, however, it is implausible to assume that the principal can form a perfect estimate. By the very nature of the regulation of risk, new and unforeseen hazards arise over time while others lose in importance. It is arguably impossible for a principal to perfectly foresee how a given agent will respond to all possible situations that might arise in the future. The principal might have access to a pool of agents whose preferences are known to be similar to her own, but there exists some residual uncertainty about the future behavior of an agent drawn from this pool.

Another motivation for the incomplete information assumption is given by the notion that the policy preferences of an agency are, at least in part, “created” by legislative mandates, agency organization, administrative procedures, and the external (interest group) environment of an agency (McCubbins et al., 1987, 1989). For example, administrative procedures influence which external interest groups have low-cost access to the agency, allowing them to bias agency decision-making to their advantage. Even if the principal designs administrative procedures to reduce “agency slack” or “political drift,” there is arguably some residual uncertainty in the workings of administrative procedures and the reactions of external interest groups that cannot be “designed away.”

We now turn to the political control problem. In practice, it is infeasible for a political principal to monitor a regulatory agency in an ongoing fashion, given the large number and technical complexity of regulatory standards. Indeed, the costs associated with perfect monitoring would more than likely undo the benefits achieved by delegating in the first place.

However, the media, interest groups, and constituents send fire-alarm signals to the principal that are informative about agency activities. We assume that the principal can costlessly receive information about the incidence of the hazards associated with the regulated activity.⁶ The “fire-alarms” in our model are not strategic players; that is, their reports are assumed to be trustworthy.

Specifically, we assume that the principal observes the agent’s decision to allow or prohibit the activity and then receives a **Good** or **Bad** fire-alarm signal, which is correlated

with the state of the world. The Bad signal occurs with probability π , the Good signal with probability $(1-\pi)$. The Bad signal implies that the state of the world is more likely to be adverse, the Good signal, favorable.⁷

Thus, the principal can distinguish the following four events. First, the agent allows the activity, after which the Bad signal is observed (aB). Second, the activity is allowed, and the Good signal is realized (aG). Third, the agent prohibits the activity, and the Bad signal occurs (pB). Fourth, the activity is prohibited, and the Good signal is realized (pG).

The primary goal of our analysis is to establish the conditions under which regulatory outcomes will be biased in the following sense. We calculate the ex ante probability that the activity is allowed given the strategic interaction between the fully informed agent and the imperfectly and incompletely informed principal; this probability is compared to the ex ante probability b that a principal who is perfectly informed about π would allow the activity. If the latter probability is lower, we define the activity to be overregulated; otherwise the activity is underregulated. Given that the agent's preferences are in effect a random variable, over- and underregulation will generally be pervasive ex post. In our view, it is of greater interest to ask whether the regulatory process is systematically biased in the ex ante sense described above.

In the model analyzed below, there are two potential sources of decision bias. First, the agent could strategically modify his decision rule to reduce the probability that his decision is overturned. Second, the principal's decision rule might not be symmetric; that is, the principal might be more likely to overturn the agent's decision if the agent allows the activity than if the activity was prohibited, or vice versa.

The time sequence of events is as follows. In period one, Nature draws the agent's type τ and the probability π that the state of the world is adverse. The adverse state of the world is realized with probability π , the favorable state with probability $(1-\pi)$. The agent privately observes τ and π . He decides whether to allow or prohibit the activity. The Bad

fire-alarm signal is realized with probability π , the Good signal with probability $(1-\pi)$. After observing the agent's decision and the fire-alarm signal, the principal decides whether to overturn the agent's decision. In period two, the adverse state of the world is realized with probability π , the favorable state with probability $(1-\pi)$. At the end of the game, the players' two-period payoffs are realized as a function of the agent's and the principal's decisions and the state of the world in each period.

The structure of the game is common knowledge. The equilibrium concept employed is that of sequential equilibrium (Kreps and Wilson, 1982).

We first analyze the decision problem of the agent. Consider the following myopic decision rule: the agent allows the activity if the probability π lies below his myopic cutpoint $b+\tau$ and prohibits it otherwise. By following this rule, the agent gets his desired policy for sure in the first period. If he modifies his strategy, he at best gets his desired policy in the second period but at the cost of having an undesired policy in the first period. By additivity of total expected utility, it follows that the agent has nothing to gain from strategically modifying his behavior. (This result would be reinforced with discounting.) Hence, the myopic decision rule is optimal.

Next, we analyze the principal's decision problem. The principal's decision to maintain or overturn the agent's decision does not affect her first-period utility. Moreover, since the agent has no active role in the second period, it does not matter whether the principal can or does dismiss the agent after the first period. For this reason, it is not the principal's primary concern to form an update on the agent's type τ . At the time of her decision, she can only influence whether the activity ends up being allowed or prohibited in the second period. (It also follows that the principal's decision is unaffected by discounting.)

When assessing her second-period payoffs, the principal's primary concern is to use the information that becomes available in the first period to form an estimate of the probability π that is constant across periods. If she estimates π to exceed b , she makes sure

that the activity is prohibited in the second period; if she estimates π to lie below b , she lets the activity be allowed. It follows that the principal overturns the agent's decision to allow the activity if her posterior expectation $E(\pi|.)$ exceeds b ; she overturns the agent's decision to prohibit the activity if $E(\pi|.)$ lies below b ; otherwise she maintains the agent's decision. (E is an expectations operator.)

The principal solves a signal extraction problem in forming an update on the probability π , based on the information conveyed by the agent's decision and the fire-alarm signal. The agent's decision depends on his type τ and the probability π , and the realization of the signal about the state of the world depends on the probability π .

Noting that there are three types of agents, Bayes' Law implies that the principal forms the update

$$E(\pi|a) = \frac{\int_0^{b-T} \pi d\pi + \int_0^b \pi d\pi + \int_0^{b+T} \pi d\pi}{(b-T) + b + (b+T)} = \frac{b}{2} + \frac{T^2}{3b} < b, \quad (3)$$

$$E(\pi|p) = \frac{\int_{b-T}^1 \pi d\pi + \int_b^1 \pi d\pi + \int_{b+T}^1 \pi d\pi}{[1-(b-T)] + (1-b) + [1-(b+T)]} = \frac{1+b}{2} - \frac{T^2}{3(1-b)} > b, \quad (4)$$

where the inequalities follows from the restriction $T \leq \min\{b, 1-b\}$.

The principal follows a no-regret override rule. Given the symmetric distribution of the agent's possible types around the principal's type, it obviously holds that $E(\pi|aG) < E(\pi|a)$ and $E(\pi|pB) > E(\pi|p)$. If the information conveyed by the fire-alarm signal appears to confirm the agent's decision--that is, when the decision to allow the hazardous activity is followed by the Good signal (aG), or the decision to prohibit the activity is followed by the Bad signal (pB)--then the principal does not overturn the agent's decision.

For the remaining outcomes, (aB) and (pG), the signal counters the agent's decision. After some tedious algebra one can show that

$$\text{sign}[b - E(\pi|aB)] = \text{sign}[b^2 - 2T^2] , \quad (5)$$

$$\text{sign}[E(\pi|pG) - b] = \text{sign}[(1-b)^2 - 2T^2] . \quad (6)$$

In interpreting (5) and (6) we distinguish two cases: the low-conflict case in which the potential conflict between the principal and her agent is weak, or T is low ($T \leq \min\left\{b\sqrt{\frac{1}{2}}, (1-b)\sqrt{\frac{1}{2}}\right\}$), and the high-conflict case in which the potential conflict is strong, or T is large ($T \geq \max\left\{b\sqrt{\frac{1}{2}}, (1-b)\sqrt{\frac{1}{2}}\right\}$).⁸

In the low-conflict case the expressions in (5) and (6) are positive so that the principal would have no incentives to overturn the agent's decision. In effect, the agent's preferences are so close to the principal's that the principal trusts the information conveyed by the agent's decision more than she trusts the information contained in the fire-alarm signal. (In the extreme case of $T=0$, the principal would obviously never have reason to overturn the decision of a clone of herself.) Some minimal conflict of interest is thus necessary to induce the principal to overturn the agent's decision. In the high-conflict case, this happens if the agent's decision to allow the activity is followed by the Bad signal, or if his decision to prohibit the activity is followed by the Good signal.

Next, we examine the existence and direction of regulatory bias. Recall that the myopic cutpoints of the three possible types, $b - T$, b , and $b + T$, are symmetric around b . Thus, the source of regulatory bias (if any) does not lie with the agent's decision rule. Of course, if the agent is of type $-T$ or T , regulatory decision-making is biased toward over- or underregulation ex post. But the agent's decision rule is not systematically biased ex ante, that is, on average across possible agent types. The first-period decision bias is zero.

In the second period, the principal never overturns the agent's decision if the potential conflict with her agent is weak. In this case, the second-period regulatory bias is also zero. Turning to the more interesting high-conflict case, it is useful to recall that the activity is

allowed in the second period if either the agent allowed the activity in the first period and the Good signal occurred (aG), in which case the agent's decision was not overturned; or if the agent prohibited the activity in the first period and the Good signal was realized (pG), in which case the principal overturned the agent's decision. The probability that the activity is allowed in the second period is thus equal to

$$\Pr(aG \cup pG) = \Pr(G) = \frac{1}{2} . \quad (7)$$

The ex ante probability that a fully informed principal would have allowed the activity is, however, equal to b . It immediately follows that there is a bias toward overregulation for $b > \frac{1}{2}$ and toward underregulation for $b < \frac{1}{2}$. Only in the "knife-edge" symmetric case of $b = \frac{1}{2}$ is there no bias.

The bias arises because the principal's choice set is discrete (override/don't override) and the binary fire-alarm signal observed by the principal is coarse relative to the continuous signal observed by the agent. The principal's no regret override rule is symmetric around one half. For $b > \frac{1}{2}$, the principal is more likely to overturn the agent's decision in the event (aB) than in the event (pG), and vice versa for $b < \frac{1}{2}$.

These two situations, $b > \frac{1}{2}$ and $b < \frac{1}{2}$, can be associated with low and high probability hazards, respectively. Maintaining the assumption of a uniform distribution, an activity can be classified as a low or high probability hazard depending on the support of the distribution of the probability π or, equivalently, the corresponding mean hazard rate Π . For expository purposes, we have considered the case where $\Pi = \frac{1}{2}$. By a simple change of units, the analysis of this case is formally equivalent to the analysis of the case in which the probability π is drawn from a uniform distribution with support on the interval from

zero to one, but b is replaced by $\frac{b}{2\Pi}$. This amounts to a scalar transformation of the players' utility functions, noting that the players are risk neutral and thus care only about means. If we measure the riskiness of a project relative to its benefits, we may say that an activity is a low probability hazard if $\Pi < b$ (equivalently, $\frac{b}{2\Pi} > \frac{1}{2}$) and a low probability hazard if $\Pi > b$ (equivalently, $\frac{b}{2\Pi} < \frac{1}{2}$). Given our normalization $\Pi = \frac{1}{2}$, it follows that the inequalities $b > \frac{1}{2}$ and $b < \frac{1}{2}$ stand for low and high probability hazards, respectively.

The model could be extended to allow for other asymmetries (for example, the number or variance of fire-alarm signals might depend on the state of the world), with similar results. In conclusion, the override model suggests that over- and underregulation will generally be pervasive and, more specifically, that low probability hazards will tend to be overregulated, high probability hazards underregulated.

3 Delegation with Dismissal

We now analyze a model in which the instrument of political oversight is dismissal. The problem at hand is an application of the classic multi-armed bandit problem [Berry and Fristedt (1985); see Banks and Sundaram (1993) for an application to elections]. After observing the agent's regulatory decision and the realization of the fire-alarm signal, the principal may choose to replace the agent with a new (random) draw from the pool of candidates. Now two types of bias can potentially arise: a decision bias would emerge if the agent strategically modified his decision rule to increase his probability of survival. (We show further below that the agent is motivated to avoid dismissal even though his preferences are defined over policy only.) In addition, a selection bias would obtain if the principal's dismissal rule implied a relatively higher survival probability for one of the extreme agents, T or $-T$, than for the other one.

The time sequence of events is modified as follows. In period one, Nature draws the first-period agent's type τ , and the probability π , that the state of the world is adverse in the

first period. The adverse state of the world is realized with probability π_1 , the favorable state with probability $(1-\pi_1)$. The agent observes τ_1 and π_1 . He decides whether to allow or prohibit the activity. The Bad fire-alarm signal is realized with probability π_1 , the Good signal with probability $(1-\pi_1)$. Based on the agent's decision and the fire-alarm signal, the principal decides whether to dismiss the agent. In period two, if the first-period agent was replaced, Nature draws the second-period agent's type τ_2 ; otherwise τ_2 is equal to τ_1 . Nature also draws the probability π_2 that the state of the world is adverse in the second period. The adverse state of the world is realized with probability π_2 , the favorable state with probability $(1-\pi_2)$. The agent observes τ_2 and π_2 . He decides whether to allow or prohibit the activity. At the end of the game, the players' two-period payoffs are realized as a function of the agent's and the principal's decisions and the state of the world in each period.

As before, the structure of the game is common knowledge, and the equilibrium concept employed is that of sequential equilibrium.

The model is solved by backwards induction. Since the second-period agent has the last move in the game, he follows his myopic optimal decision rule. Thus, the activity is allowed in the second period if the probability π_2 lies below the agent's myopic cutpoint $b + \tau_2$, and prohibited otherwise.

Turning to the principal's decision problem, it is worthwhile emphasizing an important difference between the override model and the dismissal model. In the override model, the probability that the state of the world is adverse, π , was assumed constant across periods, and the agent τ had no active role in the second period. Thus, the principal's primary concern was to form an estimate of π , with the goal of deciding whether to override the agent's decision. In the dismissal model, the second-period probability that the state of the world is adverse, π_2 , is a "fresh draw," and the second-

period agent τ_2 plays an active role. For this reason, the first-period probability π_1 , per se, is irrelevant for the principal's decision. Instead, the principal's signal extraction problem is geared towards forming an update on the first-period agent's type τ_1 based on the information conveyed by the agent's decision and the fire-alarm signal, with the goal of deciding whether to replace the first-period agent with a "fresh draw." The agent's decision depends on his type τ_1 and the probability π_1 , and the signal depends on the probability π_1 .

The principal follows a no-regret dismissal rule analogous to the no-regret override rule employed in the high-conflict case of the previous model.⁹ That is, she dismisses the agent if the fire-alarm signal counters the agent's decision, (aB) or (pG); otherwise she keeps the agent, (aG) or (pB). We show further below that this dismissal rule is optimal given the first-period agent's decision rule.

We now turn to the first-period agent's decision problem. An agent of type τ_1 employs the decision rule characterized by the cutpoint $\bar{\tau}$; this strategic cutpoint generally differs from the myopic cutpoint $b + \tau_1$. The agent trades off his first-period preferred choice with maximizing his chances of survival in order to achieve his preferred outcome in the second period. Employing the cutpoint rule $\bar{\tau}$, his probability of survival is given by

$$S(\bar{\tau}) = \int_0^{\bar{\tau}} (1-\pi) d\pi + \int_{\bar{\tau}}^1 \pi d\pi = \frac{1}{2} + \bar{\tau}(1-\bar{\tau}) . \quad (8)$$

The agent's two-period expected utility from following the cutpoint rule $\bar{\tau}$ is given by

$$E(U_1^{\tau_1} + U_2^{\tau_1} | \bar{\tau}) = \int_0^{\bar{\tau}} (b + \tau_1 - \pi) d\pi + \left[\frac{1}{2} + \bar{\tau}(1-\bar{\tau}) \right] V^{\tau_1} , \quad (9)$$

where V^{τ_1} is the value of remaining in office for an agent of type τ_1 . This continuation value is derived from the fact that an agent prefers himself or a clone of himself to make the second-period regulatory decision. If he is dismissed, the probability that he will be

replaced by an agent with different preferences is equal to two thirds. Interestingly, the agent's survival incentives are derived from his policy goals. Our results would be strengthened if we assumed that the agent cared about avoiding dismissal per se in addition to caring about the policy consequences of his dismissal.

The first-order condition for the agent's maximization problem is given by

$$\frac{\partial E(U_1^{\tau_1} + U_2^{\tau_1} | \bar{\tau})}{\partial \bar{\tau}} = b + \tau_1 - \bar{\tau} + (1 - 2\bar{\tau}) V^{\tau_1} = 0 \quad . \quad (10)$$

The term $b + \tau_1 - \bar{\tau}$ reflects the net benefits from allowing the activity in the first period, while the term $(1 - 2\bar{\tau}) V^{\tau_1}$ captures the expected benefits of survival. Eq. (10) implies that

$$\text{sign}[\bar{\tau} - (b + \tau_1)] = \text{sign}\left[\frac{1}{2} - (b + \tau_1)\right] \quad . \quad (11)$$

Intuitively, the probability of survival $S(\bar{\tau})$ peaks at $\bar{\tau} = \frac{1}{2}$. For strategic reasons, the agent's cutpoint moves away from his myopic optimum $b + \tau$, toward one half.

The two extreme agents T and $-T$ have a stronger office motivation than does the moderate agent 0 . Each extreme agent may be replaced by an extreme agent of the other kind, whose preferences are very much different from his own. In contrast, the moderate agent may be replaced by one of the extreme agents whose preferences are closer to his own. Due to the symmetric distribution of the agents' preferences, the extreme agents' survival incentives are identical in strength. In summary, the agents' continuation values are ordered as follows:

$$V^T = V^{-T} > V^0 > 0 \quad . \quad (12)$$

Since the extreme agents have identical incentives to remain in office, a bias (if any) will not be caused by an asymmetry in the agents' continuation values.

The incumbent agent in the first period now faces a tradeoff. If he follows his myopic decision rule, he maximizes his first-period utility. By strategically modifying his decision rule, he can improve his chances of survival and thereby increase his second-period expected utility. The optimal cutpoint for an agent of type τ_1 ,

$$\bar{\tau} = \frac{(b + \tau_1 + V^{\tau_1})}{1 + 2V^{\tau_1}}, \quad (13)$$

lies between the agent's myopic cutpoint and the survival maximizing cutpoint, one half:

$$b + \tau_1 > \bar{\tau} > \frac{1}{2} \quad \text{for} \quad b + \tau_1 > \frac{1}{2}, \quad (14)$$

$$b + \tau_1 < \bar{\tau} < \frac{1}{2} \quad \text{for} \quad b + \tau_1 < \frac{1}{2}. \quad (15)$$

Given the agents' equilibrium decision rules, it remains to be shown that the principal's no-regret dismissal rule is optimal. (The optimal dismissal rule is formally derived in the appendix.) Due to the symmetry of the candidates' preferences, the principal is indifferent between allowing an agent of type T or $-T$ to survive. Hence, the principal's objective is to maximize the probability that an agent of type 0 is in power in the second period. The probability that a random draw from the pool of candidates is of type 0 is equal to one third. If the posterior probability that the first-period agent is of type 0 is greater than one third, the principal will choose to keep the agent; otherwise the first-period agent is replaced with a random draw from the pool.

The principal's posterior estimates of the probability that the first-period agent is of type 0 are given by:

$$\Pr(\tau_1 = 0 | aB) < \frac{1}{3}, \quad (16)$$

$$\Pr(\tau_1 = 0 | aG) > \frac{1}{3}, \quad (17)$$

$$\Pr(\tau_1 = 0 | pB) > \frac{1}{3}, \quad (18)$$

$$\Pr(\tau_1 = 0 | pG) < \frac{1}{3} . \quad (19)$$

Thus, the principal dismisses the agent if the agent allows the activity and the Bad signal is realized (aB), or if the agent prohibits the activity and the Good signal is realized (pG); otherwise she keeps the agent.

The model is analyzed for two types of bias: decision bias and selection bias. In the first period, a decision bias arises if the average strategic cutpoint across possible agent types is not equal to b . The second-period outcome is subject to a selection bias if one of the extreme agents has a higher probability of survival than the other one.

Consider first the “knife-edge” symmetric case of $b = \frac{1}{2}$. The moderate agent’s myopic cutpoint is equal to one half, which is also the survival-maximizing cutpoint. It follows that the moderate agent’s optimal cutpoint is equal to one half [see (13)]. The strategic cutpoints of the two extreme agents are symmetric around one half [see (13)]. Thus, the first-period decision bias is zero. The probability of survival is identical for the two extreme agents [see (8)]. As a consequence, the second-period agent’s myopic cutpoint is, on average across possible types, equal to b . It follows that the second-period selection bias is also zero.

Next, consider the asymmetric case of $b > \frac{1}{2}$. The moderate agent’s myopic cutpoint is equal to b , but he strategically modifies his decision rule such that his cutpoint moves toward one half. That is, he now prohibits the activity for a larger range of realizations of π_1 in order to increase his probability of survival. The cutpoints of the extreme agents also shift toward one half. On average, the two extreme agents behave more conservatively:

$$\frac{(\bar{T} + \overline{-T})}{2} - b = \frac{(1-2b) V^{\tau_1}}{1+2V^{\tau_1}} < 0 \quad \text{for} \quad b > \frac{1}{2} . \quad (20)$$

Thus, the average cutpoint $\bar{\tau}$ across possible agent types lies below b if $b > \frac{1}{2}$: a first-period decision bias toward overregulation obtains.

Moreover, in the case of $b > \frac{1}{2}$ an agent of type $-T$ faces a more advantageous trade-off in the first period than does an agent of type T , since $-T$'s myopic cutpoint lies closer to one half. (Our statements regarding the selection bias are limited to the case in which $b - T \leq \frac{1}{2} \leq b + T$. The derivation of other cases is straightforward.) As a consequence, he has a higher survival probability than does his counterpart. On average across possible agent types, the second-period agent's myopic cutpoint lies below b in the second-period if $b > \frac{1}{2}$: there is a second-period selection bias toward overregulation.

The derivation and intuition for the case of $b < \frac{1}{2}$ follows straightforwardly. In this case, we obtain a first-period decision bias and a second-period selection bias toward underregulation.

In conclusion, the dismissal model suggests that over- and underregulation will generally be pervasive. Moreover, as before we can equate the cases $b > \frac{1}{2}$ and $b < \frac{1}{2}$ with low and high probability hazards, respectively, with the result that for low probability hazards there is a first-period decision bias and a second-period selection bias toward overregulation, whereas for high probability hazards the decision and selection biases favor underregulation.

Regulatory biases arise because the principal's choice set is discrete (dismiss/don't dismiss) and the binary fire-alarm signal observed by the principal is coarser than the continuous signal observed by the agent. The principal's no-regret dismissal rule is symmetric around one half. For $b > \frac{1}{2}$ the principal is more likely to dismiss the agent in the event (aB) than in the event (pG) ; and vice versa for $b < \frac{1}{2}$. This asymmetric political control rule leads to a selection bias: regulatory policy is biased because certain types of agents are more likely to survive. In addition, the principal's asymmetric dismissal rule

creates asymmetric incentives for different agent types to strategically modify their behavior in order to increase their probability of survival, thereby giving rise to a decision bias.

4 Discussion

Two characteristics of the political oversight process are central to our analysis: simple instruments of political oversight and coarse information flows. These features generally give rise to asymmetric decision or selection rules that lead to decision or selection biases in regulatory decision-making.

A natural extension of our model would analyze the generation of information by interested parties (Austen-Smith and Wright, 1992; Lohmann, 1993). When the policy in question is a (possibly differentiated-benefits) public good, costly communication about policy consequences is subject to a free rider problem. Policy biases might arise when the incentives to generate Good and Bad signals, respectively, are asymmetric.

Empirically, the framework explored in our paper might serve to organize our thinking about the evidence presented in informal case studies. For example, drug approval decisions made by the Food and Drug Administration (FDA) are subject to a decision bias toward overregulation because of asymmetric fire-alarm signals:

There are two types of errors the FDA can make in reviewing a new drug application: it can approve a drug that turns out to have unexpectedly adverse side effects, or it can delay or deny a beneficial drug. From a public health standpoint, both of these errors can be equally deadly, but from a political standpoint, they are worlds apart. Incorrectly approving a drug can produce highly visible victims, highly emotional news stories, and heated congressional hearings. . . . Incorrectly delaying a drug, on the other hand, will produce invisible victims and little more. . . . Not surprisingly, the FDA's fundamental approach to drug approval is designed to reduce the likelihood of the first type of error while paying little attention to the second. The well-documented result of this excessive caution is drug lag--the frequent unavailability of major new drugs . . . (Kazman, 1991, p. 5)

Kelman's (1980) analysis suggests that Occupational Safety and Hazard Administration (OSHA) decision-making is subject to a selection bias toward overregulation:

In determining the content of regulations, OSHA decision-makers have usually chosen more protective over less protective alternatives—especially in some of the more dramatic decisions the agency has made. . . . The evidence suggests that the most important factor explaining OSHA decisions on the content of regulations has been the pro-protective values of agency officials, derived from the ideology of the safety and health professional and the organizational mission of OSHA. (Kelman 1990, p. 248)

Ironically, our analysis suggests that the selection bias observed by Kelman is precisely due to the features of the political oversight process that lead him to reject the hypothesis that OSHA decision-making is shaped by political oversight:

An important alternative explanation remains: what about ‘political’ factors in OSHA’s decisions about the content of regulations? The most important barrier to White House influence over agency decision making is the fact that time is finite. Although the capacity of the president himself and of the ‘institutionalized presidency’ (the White House Staff and the Office of Management and Budget) to work endless hours is legendary, it is far outstripped by the stupendous number of important issues with which agencies deal. . . . Lack of time also means that intervention, when and if it comes, tends to come late in a drawn-out decision-making process and to be based on fragments of information rather than lengthy consideration that agency officials have given the question. It is ‘crisis-oriented.’ (Kelman, 1990, p. 250, pp. 253-54)

Clearly, the framework developed in this paper is highly exploratory and stylized. We have abstracted from the role of the tort system that provides disincentives for firms to engage in activities with excessively adverse health consequences for their workers, consumers, or the public at large. Firm liability is a close substitute for government regulation. Our analysis is partial in the sense that the overall effect of biases in regulatory policy also depends on whether they offset or exacerbate the biases inherent in an imperfect liability system.

Our analysis is incomplete in yet another respect: we do not examine possible political responses to regulatory bias. The principal could “debias” regulatory policy by designing

an employment contract that provides counteracting incentives for the agent. In political settings, however, employment contracts tend to be very simple, providing only “low-powered” incentives (Tirole, 1994). More plausibly, the principal might undo the regulatory bias by deliberately delegating authority to an agent with different preferences (Rogoff, 1985; Melumad and Mookherjee, 1989; Fershtman et al., 1991). A debiasing effect can also be achieved through procedural design. For example, the principal might make it more costly for the agent to pursue one path of action rather than another by requiring that the agent perform a costly cost-benefit analysis under certain circumstances; or by constraining the agent to follow stringent procedures in some cases, while granting him considerable discretion in others (McCubbins et al., 1987, 1989).

Endnotes

1. For simplicity, we restrict attention to the binary decision to allow or ban an activity rather than consider the more general case in which a regulator can choose the stringency of a regulatory standard.
2. The analysis could easily accommodate uncertainty about the benefits or the costs associated with the activity.
3. We make these extreme assumptions for simplicity. The following assumptions are crucial. First, the cost of setting up an agency is smaller than the difference between the expected utility achieved by the policymaker when she sets regulatory policy on the basis of her diffuse prior information and the expected utility achieved when she delegates regulatory powers to an agency. Second, once the agency is set up, the principal and the agency face different costs of information acquisition so that the agency forms a more precise estimate of the value of π .
4. The analysis can easily accommodate differences in the costs c or in the priors on π . The assumption that the pool of agents is infinite is made for technical reasons and is relevant only for the model considered in Section III, in which the agent may be dismissed: it allows us to avoid the messy calculations involved when the distribution of agents in a finite pool changes after one agent is chosen and some information about the agent's preferences is revealed through his decisions.
5. There is no loss of generality in normalizing the costs since the expected net benefits are the crucial variable driving the analysis.
6. The results would be robust if there was a small cost.
7. The assumption of discrete information flows is crucial for our analysis; the specific assumption that the fire-alarm signal is binary is made for simplicity only. The analysis also generalizes in a straightforward way to the case of multiple signals.

8. The analysis of the intermediate case in which $T \in \left(\min \left\{ b\sqrt{\frac{1}{2}}, (1-b)\sqrt{\frac{1}{2}} \right\}, \max \left\{ b\sqrt{\frac{1}{2}}, (1-b)\sqrt{\frac{1}{2}} \right\} \right)$ is more complex. In this case, the principal responds to the fire-alarm signal in an asymmetric way. For $b > \frac{1}{2}$, she overturns the agent's decision to allow the activity if the signal is Bad, whereas she never overturns the agent's decision to prohibit the activity; conversely, for $b < \frac{1}{2}$, she overturns the agent's decision to prohibit the activity if the signal is Good, whereas she never overturns the agent's decision to allow the activity. Here the source of regulatory bias is given by the principal's asymmetric response to the fire-alarm signal. This case is analyzed in Hopenhayn and Lohmann (1996).

9. Here we restrict attention to the set of $b \in [\underline{b}, (1-\underline{b})]$, $0 < \underline{b} < \frac{1}{2}$. The appendix derives other, more complex, cases involving randomized dismissal strategies. The results are qualitatively similar.

Appendix

Allowing for randomization, the principal's dismissal rule is given by the vector λ_e , $e \in \{aB, aG, pB, pG\}$, where λ_e is the probability that the agent survives (is not dismissed) if event e occurs. The following proposition summarizes the optimal dismissal rule.

Proposition. There exists a value \underline{b} , with $0 < \underline{b} < \frac{1}{2}$, such that: (i) if $b \in [\underline{b}, (1-\underline{b})]$ then the unique equilibrium is characterized by $\lambda_{aG} = \lambda_{pB} = 1$ and $\lambda_{aB} = \lambda_{pG} = 0$; (ii) if $b \in [0, \underline{b})$ then $0 < \lambda_{aG} - \lambda_{pG} < 1$ and $\lambda_{pB} = 1$, with $\lambda_{aG} - \lambda_{pG}$ uniquely determined; (iii) if $b \in (1-\underline{b}, 1]$ then $0 < \lambda_{pB} - \lambda_{aB} < 1$ and $\lambda_{aG} = 1$, with $\lambda_{pB} - \lambda_{aB}$ uniquely determined.

Proof of Proposition. The proof of the proposition involves the following steps. (i) Establish that the agents' decision rules (allow/prohibit) are characterized by cutpoint values $\bar{\tau}$, $\tau_i \in \{-T, 0, T\}$. (ii) Show that the corresponding cutpoints satisfy $-\bar{T} < \bar{0} < \bar{T}$. (iii) Derive the properties given in the proposition.

Lemma 1. There exist cutpoints $\bar{\tau}$, $\tau_i \in \{-T, 0, T\}$, such that agent τ allows the activity *iff* $\pi_i \leq \bar{\tau}$. Furthermore, $-\bar{T} < \bar{0} < \bar{T}$.

Proof of Lemma 1. For agent τ_i , the difference between the expected benefits of allowing the activity and the expected benefits of prohibiting the activity are given by

$$G(\tau_i, \pi_i) \equiv b + \tau_i + (\lambda_{aG} - \lambda_{pG}) V^{\tau_i} - \pi_i [1 + (\lambda_{aG} + \lambda_{pB} - \lambda_{aB} - \lambda_{pG}) V^{\tau_i}] .$$

It is easy to verify, first, that $V^{\tau_1} < G(\tau_1, \pi_1) < \frac{1}{2}$ so that $G(\tau_1, \pi_1)$ is strictly decreasing in π_1 , and, second, that $G(\tau_1, 0) > 0$ and $G(\tau_1, 1) < 0$. Let $\bar{\tau}$ be the unique value such that $G(\tau_1, \bar{\tau}) = 0$. Agent τ_1 allows the activity iff $\pi_1 \leq \bar{\tau}$. From the definition of $G(\tau_1, \pi_1)$ it is straightforward to verify that $\bar{\tau}$ is increasing in $b + \tau_1$.

The following lemma characterizes some properties of the posterior distributions on the agent's type. We develop only one of the posterior distributions here since the remaining ones follow immediately. The posterior $\Pr(\tau_1 | aB)$ satisfies

$$\Pr(\tau_1 = 0 | aB) = \frac{\int_0^{\bar{0}} \pi d\pi}{\int_0^{\bar{0}} \pi d\pi + \int_0^{-\bar{\tau}} \pi d\pi + \int_0^{\bar{\tau}} \pi d\pi} = \frac{\int_0^{\bar{0}} \pi d\pi}{3 \int_0^{\bar{0}} \pi d\pi + \int_0^{\bar{\tau}} \pi d\pi - \int_0^{-\bar{\tau}} \pi d\pi} \quad (A1)$$

The agent is not removed iff $\Pr(\tau_1 = 0 | aB) \geq \frac{1}{3}$, the unconditional probability of type 0 in the pool of agents. Eq. (A1) immediately implies that this inequality holds iff $\int_0^{\bar{\tau}} \pi d\pi \leq \int_0^{-\bar{\tau}} \pi d\pi$. Applying a similar argument, the following lemma is proved.

Lemma 2. The posterior probabilities satisfy the following inequalities:

- (i) $\Pr(\tau_1 = 0 | aB) \geq \frac{1}{3}$ iff $\int_0^{\bar{\tau}} \pi d\pi \leq \int_0^{-\bar{\tau}} \pi d\pi$,
- (ii) $\Pr(\tau_1 = 0 | aG) \geq \frac{1}{3}$ iff $\int_0^{\bar{\tau}} (1 - \pi) d\pi \leq \int_0^{-\bar{\tau}} (1 - \pi) d\pi$,
- (iii) $\Pr(\tau_1 = 0 | pB) \geq \frac{1}{3}$ iff $\int_0^{\bar{\tau}} \pi d\pi \geq \int_0^{-\bar{\tau}} \pi d\pi$,

$$(iv) \quad \Pr(\tau_1=0 | pG) \geq \frac{1}{3} \quad \text{iff} \quad \int_{\bar{0}}^{\bar{r}} (1-\pi) d\pi \geq \int_{-\bar{r}}^{\bar{0}} (1-\pi) d\pi .$$

Corollary. (i) $\Pr(\tau_1=0 | aB) < \frac{1}{3} \Leftrightarrow \Pr(\tau_1=0 | pB) > \frac{1}{3}$, and $\Pr(\tau_1=0 | aB) = \frac{1}{3} \Leftrightarrow \Pr(\tau_1=0 | pB) = \frac{1}{3}$; similar properties hold for $\Pr(\tau_1=0 | aG)$ and $\Pr(\tau_1=0 | pG)$; (ii) $\Pr(\tau_1=0 | aB) \geq \frac{1}{3} \Rightarrow \Pr(\tau_1=0 | aG) > \frac{1}{3}$; (iii) $\Pr(\tau_1=0 | pG) \geq \frac{1}{3} \Rightarrow \Pr(\tau_1=0 | pB) > \frac{1}{3}$.

Lemma 2 and the Corollary significantly restrict the type of equilibria that can arise. These restrictions are summarized in the following lemma.

Lemma 3. The equilibrium dismissal rule satisfies: (i) if $\lambda_{aG} < 1$, then $\lambda_{aB} = 0$ and $\lambda_{pB} = 1$; (ii) if $\lambda_{pB} < 1$, then $\lambda_{pG} = 0$ and $\lambda_{aG} = 1$; (iii) if $\lambda_{pG} > 0$, then $\lambda_{aB} = 0$, and, conversely, if $\lambda_{aB} > 0$, then $\lambda_{pG} = 0$.

Proof of Lemma 3. The lemma is an immediate consequence of the Corollary.

Note that Lemma 3 implies that at least one of the no-dismissal probabilities ($\lambda_{aG}, \lambda_{pB}$) is equal to one, while at least one of the no-dismissal probabilities ($\lambda_{aB}, \lambda_{pG}$) is zero. A comprehensive and exclusive list of candidate equilibria is given by:

- (a) $\lambda_{aG} = \lambda_{pB} = 1; \lambda_{aB} = \lambda_{pG} = 0;$
- (b) $\lambda_{aG} = \lambda_{pB} = 1; \lambda_{aB} > 0; \lambda_{pG} = 0;$
- (c) $\lambda_{aG} = \lambda_{pB} = 1; \lambda_{pG} > 0; \lambda_{aB} = 0;$
- (d) $\lambda_{aG} = 1; \lambda_{pB} < 1; \lambda_{pG} = 0; \lambda_{aB} \geq 0;$

(e) $\lambda_{aG} < 1; \lambda_{pB} = 1; \lambda_{aB} = 0; \lambda_{pG} \geq 0.$

The first item in the list is the only possible equilibrium that does not involve randomization. The unique equilibrium is of this type provided b is not too far from one half. Otherwise, the equilibrium is mixed, and a source of indeterminacy arises, which we now discuss.

Suppose there exists an equilibrium characterized by $\int_0^{\bar{\tau}} (1-\pi) d\pi = \int_{-\bar{\tau}}^{\bar{0}} (1-\pi) d\pi.$

Since this implies that $\Pr(\tau_1=0 | aG) = \Pr(\tau_1=0 | pG) = \frac{1}{3}$, λ_{aG} and λ_{pG} are not restricted. The cutpoint rule $\bar{\tau}$ satisfies

$$\bar{\tau} = \frac{b + \tau_1 + (\lambda_{aG} - \lambda_{pG}) V^{\tau_1}}{1 + (\lambda_{aG} - \lambda_{pG} + \lambda_{pB} - \lambda_{aB}) V^{\tau_1}}, \quad (A2)$$

where (A2) must hold for $\tau_1 \in \{-T, 0, T\}$. This equation restricts $(\lambda_{aG} - \lambda_{pG})$ but not the individual components, hence an indeterminacy arises. Similar remarks can be made for

λ_{pB} and λ_{aB} when $\int_0^{\bar{\tau}} \pi d\pi = \int_{-\bar{\tau}}^{\bar{0}} \pi d\pi.$

We now establish that aside from this indeterminacy the equilibrium is unique and that there exist well-defined bounds for b that define the regions separating the different types of equilibrium described above.

The condition $\int_0^{\bar{\tau}} (1-\pi) d\pi = \int_{-\bar{\tau}}^{\bar{0}} (1-\pi) d\pi$ can be rewritten as

$$\frac{\bar{T} \left(1 - \frac{\bar{T}}{2}\right) + -\bar{T} \left(1 - \frac{-\bar{T}}{2}\right)}{2} = \bar{0} \left(1 - \frac{\bar{0}}{2}\right). \quad (A3)$$

To analyze equilibria of type (c) and (e), define $R(x) \equiv x \left(1 - \frac{x}{2}\right)$, $\pi(\lambda, V, b) \equiv \frac{b + \lambda V}{1 + (1 + \lambda)V}$, and $W(\lambda, V, b) \equiv R[\pi(\lambda, V, b)]$. Letting $\lambda \equiv (\lambda_{aG} - \lambda_{pG})$, noting that $\lambda_{pB} = 1$ and $\lambda_{aB} = 0$, and using (A2), after some calculations we can rewrite (A3) as follows:

$$W(\lambda, V^0, b) - W(\lambda, V^T, b) + \frac{T^2}{2 [1 + (1 + \lambda) V^T]^2} = 0 \quad (\text{A4})$$

Since $V^0 < V^T$, a necessary condition for this equality to hold is that $\frac{\partial W}{\partial V} > 0$. But since $\frac{\partial R}{\partial \pi} = 1 - \pi > 0$, this condition is equivalent to $0 < \text{sign} \left(\frac{\partial \pi}{\partial V} \right) = \text{sign}[\lambda - b(1 + \lambda)]$. For this inequality to hold, it must be the case that $b < \frac{\lambda}{1 + \lambda}$, which in particular implies that $b < \frac{1}{2}$.

The following lemma will be used to establish the conditions under which (A4) is satisfied.

Lemma 4. $W_{12} > 0$ and, for $b \leq \frac{1}{2}$, $W_{23} < 0$.

Proof of Lemma 4. The proof involves tedious calculations and is available from the authors upon request.

Lemma 4 implies that the left hand side of (A4) is strictly increasing in λ and, in particular, that this expression is maximized at $\lambda = 1$ and minimized at $\lambda = 0$. A straightforward calculation shows that this expression is negative at $\lambda = 0$. Lemma 4 also implies that the left hand side of (A4) is strictly decreasing in b . It follows that this

expression is maximized at $b=0$. Lengthy and tedious calculations establish that at $\lambda=1$ (A4) becomes positive as b goes to zero. Using these results, we may prove:

Lemma 5. For each $G>0$, there exists a value $b(G)$, with $0<b(G)<\frac{1}{2}$, such that: (i) if $b\in [b(G),(1-b(G))]$ then the unique equilibrium is of type (a); (ii) if $b\in [0,b(G))$ then the equilibria are of type (d) and (e), depending on the specific choices of λ_{aG} and λ_{pG} , with $\lambda_{aG}-\lambda_{pG}>0$ uniquely determined; (iii) if $b\in ((1-b(G)),1]$ then the equilibria are of type (b) and (d) depending on the specific choices of λ_{pB} and λ_{aB} , with $\lambda_{pB}-\lambda_{aB}>0$ uniquely determined.

Proof of Lemma 5. Parts (i) and (ii) follow directly from the above remarks. Part (iii) is

obtained by noting that $1-\bar{\tau} = 1 - \frac{b + \tau_1 + (\lambda_{pB} - \lambda_{aB}) V^{\tau_1}}{1 + (\lambda_{aG} - \lambda_{pG} + \lambda_{pB} - \lambda_{aB}) V^{\tau_1}}$ so that $1-\bar{\tau}$ and $1-$

$(b+\tau_j)$ take the place of $\bar{\tau}$ and $b+\tau_j$ in the above derivation.

References

- Austen-Smith, D., and Wright, J. (1992). "Competitive Lobbying for a Legislator's Vote." Social Choice and Welfare **9**: 229-257.
- Banks, J.S., and Sundaram, R.K. (1993). "Adverse Selection and Moral Hazard in a Repeated Elections Model," in Political Economy: Institutions, Information, Competition and Representation (W. Barnett, M. Hinich, and N. Schofield, Eds.), Ch. 12. Cambridge: Cambridge University Press.
- Berry, D.A., and Fristedt, B. (1985). Bandit Problems: Sequential Allocation of Experiments. London: Chapman and Hall.
- Fershtman, C., Judd, K.L., and Kalai, E. (1991). "Observable Contracts: Strategic Delegation and Cooperation," International Economic Review **32**: 551-560.
- Hopenhayn, H., and Lohmann, S. (1996). "Fire-Alarm Signals and the Political Oversight of Regulatory Agencies," Journal of Law, Economics, and Organization **12**: 199-216.
- Kahnemann, D., Slovic, P., and Tversky, A., Eds. (1982). Judgement Under Uncertainty: Heuristics and Biases. Cambridge: Cambridge University Press.
- Kazman, S. (1991). "Death by Regulation," Regulation: The Cato Review of Business and Government **14**: 18-22.
- Kelman, S. (1980). "Occupational Safety and Health Administration," in The Politics of Regulation (J. Q. Wilson, Ed.), Ch. 7. New York: Basic Books.
- Kiewiet, D.R., and McCubbins, M.D. (1991). The Logic of Delegation. Chicago: University of Chicago Press.
- Kreps, D.M., and Wilson, R. (1982). "Sequential Equilibria," Econometrica **50**: 863-893.
- Lohmann, S. (1993). "A Signaling Model of Informative and Manipulative Political Action," American Political Science Review **87**: 319-333.
- Lupia, A. (1992). "Busy Voters, Agenda Control, and the Power of Information," American Political Science Review **86**: 390-403.
- Lupia, A. (1994). "Shortcuts Versus Encyclopedias: Information and Voting Behavior in California Insurance Reform Elections," American Political Science Review **88**: 63-76.
- Magat, W.A., Krupnik, A.J., and Harrington, W. (1986). Rules in the Making: A Statistical Analysis of Agency Behavior. Washington, D.C.: Resources for the Future.
- McCubbins, M.D., Noll, R.G., and Weingast, B.R. (1987). "Administrative Procedures as Instruments of Political Control," Journal of Law, Economics, and Organization **3**: 243-278.

- McCubbins, M.D., Noll, R.G., and Weingast, B.R. (1989). "Structure and Process: Politics and Policy: Administrative Arrangements and the Political Control of Agencies," Virginia Law Review **75**: 431-482.
- McCubbins, M.D., and Schwartz, T. (1984). "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms," American Political Science Review **28**: 165-179.
- Melumad, N.D., and Mookherjee, D. (1989). "Delegation as Commitments: The Case of Income Tax Audits," RAND Journal of Economics **20**: 139-63.
- Office of Management and Budget (1992). President's Budget for Fiscal Year 1992, Ch. 1X.C. Washington, D.C.
- Rogoff, K. (1985). "The Optimal Degree of Commitment to an Intermediate Monetary Target," Quarterly Journal of Economics **100**: 1169-1189.
- Slovic, P., Fischhoff, B., and Lichtenstein, S. (1985). "Regulation of Risk: A Psychological Perspective," in Regulatory Policy and the Social Sciences (R.G. Noll, Ed.). Berkeley: University of California Press.
- Tirole, J. (1994). "The Internal Organization of Government," Oxford Economic Papers **46**: 1-29.
- Viscusi, W.K. (1992). Fatal Tradeoffs: Public and Private Responsibilities for Risk. New York: Oxford University Press.
- Wildavsky, A. (1995). But Is It True? A Citizen's Guide to Environmental Health and Safety Issues. Cambridge: Harvard University Press.

List of Symbols

U^P	principal's one-period utility
b	benefits of activity if allowed
c	costs of activity if allowed and state of world is adverse
π	probability that state of world is adverse
Π	mean probability that state of world is adverse
U^τ	one-period utility of agent of type τ
$\tau \in \{T, 0, -T\}$	agent type τ can take on one of three values: T , 0 , or $-T$
T	parameter capturing strength of conflict of interest between principal and agent
(aB)	event (agent a llows activity and B ad signal is realized)
(aG)	event (agent a llows activity and G ood signal is realized)
(pB)	event (agent p rohibits activity and B ad signal is realized)
(pG)	event (agent p rohibits activity and G ood signal is realized)
E	expectations operator
subscript 1	indexes first-period values
subscript 2	indexes second-period values
	example:
$U_2^{\tau_1}$	second-period utility (U_2) of agent appointed in first period (τ_1)
$\bar{\tau} \in \{\bar{T}, \bar{0}, \bar{-T}\}$	allow/prohibit cutpoint for agent of type τ_1 : takes on value \bar{T} for agent of type T , $\bar{0}$ for agent of type 0 , and $\bar{-T}$ for agent of type $-T$
S	probability that agent survives (does <u>not</u> get dismissed)
$V^{\tau_1} \in \{V^T, V^0, V^{-T}\}$	continuation value for agent of type τ_1 ; takes on value V^T for agent of type T , V^0 for agent of type 0 , and V^{-T} for agent of type $-T$

Biographical Sketch

Susanne Lohmann is a Visiting Associate Professor of Political Science at the California Institute of Technology and an Associate Professor of Political Science at the University of California, Los Angeles. Her articles on collective action, central banking, and international political economy have appeared in American Economic Review, American Political Science Review, Economics and Politics, European Economic Review, International Organization, Journal of Law, Economics, and Organization, Public Choice, World Politics, and other journals. She is currently working on a book about the rationality of mass action, with applications to the East German revolution and German unification.