

# Clarity of Responsibility with Sequential Policy Making\*

Ethan Bueno de Mesquita<sup>†</sup>

Dimitri Landa<sup>‡</sup>

January 11, 2012

## Abstract

We analyze the effects of clarity of responsibility when policy making is sequential, showing that the two components of clarity of responsibility are negative complements: taken together, unified agency and observability of agent actions can reduce the welfare of political principals. In equilibrium, the principal engages in “What Have You Done for Me Lately” (WHYDFML) behavior—inducing the agent to over-emphasize the late stages of the policy-making process relative to early stages. Eliminating either component of clarity of responsibility eliminates WHYDFML-driven inefficiencies, but at a cost in terms of the power of incentives and level of rent seeking. Whether complete clarity of responsibility or some more limited form of it is optimal depends on the magnitudes of these competing effects. Although eliminating both aspects of clarity of responsibility is never optimal, it is possible for no clarity of responsibility to dominate complete clarity of responsibility.

---

\*We benefitted from conversations with Scott Ashworth, Randy Calvert, Eric Dickson, Amanda Friedenber, Sean Gilmard, Jake Gersen, Sandy Gordon, Cathy Hafer, Alberto Simpser, David Stasavage, Matthew Stephenson, Alan Wiseman, and seminar participants at the University of Chicago. This research was supported by NSF grants SES-0819152 (Bueno de Mesquita) and SES-0819545 (Landa)

<sup>†</sup>Professor, Harris School of Public Policy Studies, University of Chicago, e-mail: bdm@uchicago.edu.

<sup>‡</sup>Associate Professor, Wilf Family Department of Politics, NYU, e-mail: dimitri.landa@nyu.edu

Accountability in any form of political representation hinges on the principals' ability to tailor sanctions and rewards to choices made by their agents. That ability depends, in turn, on the principals' access to a clear picture of those choices and of their relationship to the outcomes they seek to influence. The concept of *clarity of responsibility*, which has gained considerable influence in comparative political analysis, is an immediate outgrowth of this view (Lewis-Beck, 1988; Powell and Whitten, 1993; Powell, 2000).

Clarity of responsibility is understood to obtain when a principal can clearly identify the agents and actions responsible for particular outcomes. It is, thus, affected by two features of the institutional environment: the number of agents and the observability of the agents' actions. Consolidating control in a single agent helps induce clarity of responsibility by minimizing the ability of agents to shift the blame for policy failures onto other agents. Making agent actions observable helps induce clarity of responsibility by making it possible to determine the extent to which particular agents' actions, rather than stochastic factors or other agents' actions, led to policy success or failure.

We develop a model of clarity of responsibility that incorporates both of these institutional features. The model captures existing intuitions about the benefits of clarity of responsibility, but it also shows them to be critically incomplete when the policy-making process is sequential. Indeed, we uncover effects of clarity of responsibility that can sometimes change the normative conclusions regarding its ultimate desirability.

We show that when the policy-making process is sequential, introducing either institutional feature of clarity of responsibility on its own improves the principal's welfare. However, introducing these two institutional features in tandem need not do so. The reason is that when both action observability and unified agency are in place, the principal engages in "What Have You Done for Me Lately" (WHYDFML) behavior in equilibrium—inducing the agent to allocate too much effort to the late stages of the policy-making process relative to early stages.<sup>1</sup>

The fact that clarity of responsibility induces WHYDFML behavior creates a trade-off in evaluating the optimal institutional design. Eliminating action observability or unified agency eliminates WHYDFML. However, without action observability, the principal is not able to give the agent as powerful incentives, and without unified agency each agent controls fewer resources, making rent seeking more attractive. Whether complete clarity of responsibility or some more limited form of it

---

<sup>1</sup>Weingast, Shepsle and Johnsen (1981) coined the term WHYDFML.

is optimal depends on the details of the institutional environment, which affect the relative magnitudes of these competing effects. We find that eliminating both aspects of clarity of responsibility is never optimal, since doing so introduces precisely the team production problem among the agents that motivates defenders of clarity of responsibility. That said, we also show that sometimes having no clarity of responsibility (i.e., neither unified agency nor action observability) is better for the principal than having complete clarity of responsibility.

The sequential policy-making process that underlies our results is empirically common, even if its analysis is not. (See Stephenson (2008) for examples.) In regulatory politics, new regulations are drafted and then implemented. Each of these steps involves costly effort which may be taken by the same or different agencies or departments within an agency. Intelligence agencies must first gather and then analyze information. Recent policy debates in and around the intelligence community demonstrate that there is considerable disagreement regarding the value of separation or integration of these functions within a single agency. Yet another example is law enforcement, where investigations necessarily precede prosecutions. In some settings this sequential process is carried out by separate agencies (e.g., the United States where in many jurisdictions there are separate elected sheriffs and district attorneys) while in others both tasks are performed by one integrated agency (e.g., French magistrates).

Our analysis of clarity of responsibility yields both substantive and methodological implications for our understanding of principal-agent relationships in politics.

First, it provides new insights for two on-going debates on the design of political institutions: one exploring the consequences of increasing the transparency of the policy-making process,<sup>2</sup> and the other exploring the consequences of “unbundling” executive powers by devolving them to multiple agents.<sup>3</sup> Our model incorporates both transparency (modeled as action observability) and unified agency as institutional variables, showing the fundamental relationship between the incentives each of them creates and identifying wholly new mechanisms for the potential virtues of decreased

---

<sup>2</sup>See, for example, Canes-Wrone, Herron and Shotts (2001); Maskin and Tirole (2004); Fox (2007); Fox and Shotts (2009); Ashworth and Shotts (2010); Shotts and Wiseman (2010); Fox and Stephenson (2011), and Fox and van Weelden (2012), focusing on how transparency creates incentives for pandering and other forms of political posturing.

<sup>3</sup>See, for example, Besley and Coate (2003); Berry and Gersen (2008), and Gersen (2010), focusing primarily on capture by special interests and the costs of oversight.

transparency or unbundled executive authority.

Second, our findings suggest subtle ways in which the accountability-maximizing institutional design of an agency may depend on the particular mission of that agency. For instance, we argue that clarity of responsibility may promote accountability in agencies, like the SEC, where actions later in the regulatory process (e.g., monitoring and enforcement) are most important, but may diminish accountability in agencies, like the EPA, where actions early in the regulatory process (e.g., drafting technical rules) carry greater weight.

Third, our model speaks to the optimal design of policy given institutional constraints. We identify ways in which policy makers ought to take the WHYDFML effect into account when crafting policies to be implemented by agencies with the institutional properties we analyze.

Finally, we make a methodological point. An implicit assumption in much of the formal literature on political agency is that, when there is only one policy output, it suffices to consider the agent taking only one action. We show that models with one observable action may not be good approximations of situations with multiple observable actions, underscoring the importance of explicitly modeling the sequential nature of many real-world policy-making processes.

## 1 The Basic Story

To get a sense of how our model works, consider the following heuristic story. A political principal is attempting to influence the behavior of a regulatory agency (the agent) with respect to some policy. The political principal might be a lobbyist, an interest group, or some other interested party. Importantly for this example, suppose that it is normatively desirable for the principal to wield this influence. The policy will be made by the agency in two stages: first regulations will be designed and then they will be enforced. The only leverage the principal has with the agency is the threat to withdraw its public support for the agency and its policies.

During the first stage of the process, while the regulation is being drafted, the principal wants to use the threat of withdrawing her support to induce the agency to adopt favorable regulations. But she may not be able to do so credibly. The principal's credibility problem emerges because of the second stage of the policy-making process—the enforcement stage. Even if the principal is unhappy with the regulation as drafted, she will not withdraw her support prior to the enforcement stage

since, if she were to do so, she would lose her leverage over the agency's enforcement decisions—decisions which she would want to influence even while being unhappy with the regulations. The agency, anticipating this fact, will only be responsive to the principal in the later stages of the policy-making process.

Hence, the sequential nature of the policy-making process limits the amount of control the principal can wield at the policy design stage. Indeed, she can only wield influence at that stage insofar as following through on threats made at the policy design stage is consistent with the principal also achieving her goals at the enforcement stage. If not, then once the enforcement stage arrives, the principal will have an incentive to renege on her threat to punish the agency for designing a regulation the principal doesn't like. As a result, the principal ends up giving the agency stronger incentives late in the policy-making process.

There are a couple of institutional ways that the principal might address this inefficiency. First, the principal could divide control over the two stages of the policy-making process between two different agencies. Second, the principal could leave authority unified, but give up on separately observing the two parts of the process—instead just observing some summary statistic such as the success or failure of the overall policy initiative. Each of these institutional shifts away from full clarity of responsibility prevents the principal from providing the kind of distorted incentives described above. However, each also comes at a cost in terms of the power of the incentives the principal can give the agent.

## **2 Clarity of Responsibility and Related Concepts**

Lewis-Beck (1988) and Powell and Whitten (1993) are the initial articulations of the argument that institutions that make clear each agent's contribution to the policy outcome promote accountability. The extensive empirical literature that followed includes Whitten and Palmer (1999), Powell (2000), Royed, Leyden and Borrelli (2000), Nadeau, Niemi and Yoshinaka (2002), Samuels (2004), Duch and Stevenson (2007), Tavits (2007), and others.

The comparative politics literature has focused on advancing clarity of responsibility by exercising two institutional levers: unified agency and action observability. Focusing on the first, Tavits (2007, p. 221) writes that “when a single party occupies the main offices of the executive branch

and possesses control over a parliament that initiates and changes policies, citizens are provided with maximum clarity of responsibility.” Focusing on the second, Alt and Lassen (2006, p. 1406) maintain that “greater transparency eases the task of attributing outcomes to the acts of particular politicians. It makes observers more able to distinguish effort from opportunistic behavior or stochastic factors. . . the political science literature calls this consequence of transparent institutions ‘clarity of responsibility’.” Given this, we reserve the term *clarity of responsibility* for institutional environments with both unified agency and action observability. Our analysis then isolates the effects of each of these institutional features.

The concept of clarity of responsibility is related to ideas of team production and multitasking in industrial organization (Holmström, 1982; Holmström and Milgrom, 1991). Our model is also connected to the literature on sequential multitask in principal-agent models (see, for example, Riordan and Sappington, 1987; Laffont and Tirole, 1988; Baron and Besanko, 1992; Gilbert and Riordan, 1995; Lewis and Sappington, 1997; Khalil, Kim and Shin, 2006). However, those models assume principals can write the type of contracts that are standard in economic environments. As is appropriate for studying political environments, our principals have a more limited set of rewards and punishments available (in particular, whether or not to retain the agent) and so behavior is quite different from that found in industrial organization.

Within the literature on political institutions, our model relates to and differs from analyses of two further concepts: separation of powers and bureaucratic redundancy. Separation of powers suggests at least some diffusion of responsibility, but may also entail other features, such as specific division of tasks and various mechanisms of checks and balances (Persson, Roland and Tabellini, 1997; Gailmard and Patty, 2009). For this reason, separation of powers is not synonymous with lack of clarity of responsibility. Like clarity of responsibility, bureaucratic redundancy also concerns the effects of unified agency (Ting, 2003). However, a key feature of redundancy is that agents are involved in independent rather than joint production, whereas the key argument for clarity of responsibility is precisely that policy production is joint.<sup>4</sup>

---

<sup>4</sup>This also distinguishes our work from Hatfield and Padró i Miquel (2006), which considers a model of multi-tasking in which each allocation contributes to a separate public good.

### 3 The Model

Ours is a moral hazard model of political agency, in the spirit of Barro (1973), Ferejohn (1986), Austen-Smith and Banks (1989), Seabright (1996), Persson, Roland and Tabellini (1997), Persson and Tabellini (2000), Shi and Svensson (2006), and others. The analysis of such models has typically focused on equilibria in which the principal uses a retention rule that, when best responded to by the agent, maximizes the principal's welfare. The idea, made explicit in Persson and Tabellini (2000), is that before the agent acts, the principal will communicate the retention rule she intends to use. The agent will find this communication credible because the principal's choice at the point of retention has no effect on future play. We want to analyze the effects of clarity of responsibility in situations where actions are taken sequentially. As such, we must consider the possibility that the principal will want to alter her stated retention rule between the actions of the agent. Agents, of course, will take this possibility into account when making their choices. The strategic nature of such interactions underscores the value of modeling communication from the principal to the agent explicitly, and we do so below.

#### 3.1 Actors and Order of Play

In all games we consider, there is one principal. The games differ with respect to two factors. First, the two actions that influence the outcome may be chosen by either one or two agents (unified versus divided agency). Second, the principal may observe these actions directly or observe only the final policy outcome (action observability versus no action observability).

Play is comprised of two kinds of stages: *policy-making* and *retention* stages. The game starts with two policy-making stages. At the beginning of policy-making stage  $t$ , the principal sends a cheap-talk message  $m_t$ , interpreted as the principal's declaration of her retention rule(s). The agent(s) observe the message  $m_t$  and the agent charged to act in policy-making stage  $t$  takes action  $a_t$ . (If there is one agent, he acts in both policy-making stages. If there are two agents, one acts in policy-making stage 1 and the other in policy-making stage 2.) The game ends with the retention stage. In the retention stage, the principal chooses to retain or dismiss each agent according to the rule specified by  $\rho$ .

In all games, the total budget available for the two actions is  $\bar{a}$ . When there is unified agency,

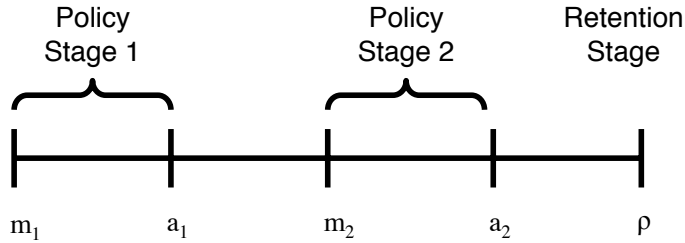


Figure 1: Timeline of games.

this budget is fungible between the two tasks. When there is divided agency, the budget available for each task is  $\frac{\bar{a}}{2}$ . We interpret the actions as resource allocations. A timeline is represented in Figure 1.

The resource allocations influence whether the final policy outcome is a success ( $s$ ) or a failure ( $f$ ). The probability that the policy succeeds is  $p(a_1, a_2)$ . We assume that  $p(\cdot, \cdot)$  is increasing, concave, and continuously differentiable in each of its arguments. Further, we assume a set of Inada conditions on  $p(\cdot, \cdot)$ , but defer their introduction until the subsection on the players' payoffs.

A retention rule maps from allocation choices and outcomes into a probability of retention. Slightly abusing notation, in each of the games we will denote the set of all possible retention rules  $\mathcal{R}$ . Retention decisions are made according to the retention rule used in the retention stage, regardless of the principal's message(s) in the policy-making stages. That is, the retention rule(s) declared during the policy-making stages are cheap talk.

### 3.2 Strategies

A strategy for an agent (referred to as “he”) is a mapping from histories into a choice of action (or actions). A strategy for the principal (referred to as “she”) is a message choice in the first policy-making stage, a history-contingent message choice in the second policy-making stage, and a history-contingent choice of retention rule in the retention stage. If there are two agents, then each element of the principal's strategy involves a retention rule for each agent.

### 3.3 Payoffs

The principal cares only about the success of the policy. We assume the principal has a von Neumann-Morgenstern expected utility function and, without loss of generality, normalize the

payoff from success to one and the payoff from failure to zero.

Agents value two things: retention and rents from resources not allocated to policy making. Retention is more valuable when there is only one agent (though the two values of retention may be arbitrarily close to equal). In particular, we assume that when there is one agent the benefit of retention is  $\bar{B}$  and when there are two agents the benefit of retention is  $\underline{B} < \bar{B}$ .

The function  $u(\cdot)$  represents an agent's payoffs from resources that he controls but does not expend on policy (i.e., rent seeking). We assume that  $u$  is increasing and weakly concave. An agent  $i$  has preferences given by the following von Neumann-Morgenstern expected utility function:

$$v_i(A) = \begin{cases} B + u(A) & \text{if } i \text{ is retained} \\ u(A) & \text{else,} \end{cases}$$

where  $A$  represents the resources  $i$  controls but does not allocate to policy.

The following Inada conditions insure interior solutions in games without action observability:  $p_1(\frac{\bar{a}}{2}, a_2)\bar{B} < u'(0)$  for all  $a_2$ ,  $p_2(a_1, \frac{\bar{a}}{2})\bar{B} < u'(0)$  for all  $a_1$ , and  $\lim_{a_i \rightarrow 0} p_i(a_1, a_2) = \infty$ , for  $i = 1, 2$ .

### 3.4 Solution Concept

Our basic solution concept is subgame perfect Nash equilibrium (SPNE). Since declarations of retention rules during policy-making stages are cheap talk, SPNE does not prevent the principal from declaring a retention rule early in the game and making a different declaration or using a different rule later. Moreover, the retention rule actually used is chosen after all actions are taken, so any rule is sequentially rational. As is usual in such an environment, there are many equilibria.

Our selection criterion is a generalization to the sequential policy-making environment of the standard criterion based on the principal's welfare. (As will be clear the application of our selection criterion to a game without sequential policy-making would select the standard, principal-welfare maximizing equilibrium.) As we noted above, that criterion is typically justified by reference to non-modeled communication from the principal. Given our explicit modeling of that communication, our approach is as follows. We put some structure on the relationship between declarations by the principal and the final retention rule applied by slightly relaxing the assumption that declarations by the principal are pure cheap talk. The intuition is that the principal may face small costs—

reputational, psychological, or in terms of communication—for changing her mind. As such, we select those equilibria in our pure cheap-talk game that are robust to the presence of small costs for the principal if she revises a declared retention rule from one stage to another.

More formally, for any of our games, consider a perturbed game in which the principal bears a cost,  $\epsilon > 0$ , for each declared retention rule that she changes from one stage to the next. We will call an SPNE *robust to small revision costs* if it is the limit of a sequence of equilibria in a sequence of nearby games as the revision cost,  $\epsilon$ , goes to zero. A formal definition is provided at the beginning of Appendix A. We refer to an SPNE that is robust to small revision costs as an *equilibrium*.

## 4 Benchmark: No Clarity of Responsibility

We begin with the benchmark case of no clarity of responsibility—divided agency and no action observability. A strategy for the principal in this game is a triple,  $(m_1, m_2, \rho)$ . Since there are two agents, each element of the triple is composed of two components (i.e., a retention rule for each agent). We write these retention rules only as mappings from outcomes into retention probabilities, since they must be constant in action. The retention rules, thus, will take only two values: a retention probability conditional on success and a retention probability conditional on failure.

A strategy for agent 1 is a choice,  $a_1(m_1)$ , of an action given an initially declared retention rule, and a strategy for agent 2 is a choice,  $a_2(m_1, m_2)$ , of an action given both declarations.

The basic intuition for equilibrium behavior is as follows. First, consider the end of the game. At the retention stage, all agent actions have already been taken. Thus, for any positive revision cost,  $\epsilon > 0$ , the principal will apply whatever retention rules she announced during the second policy-making stage. Anticipating this, in the second policy-making stage, the relevant agent will best respond to whatever retention rule the principal just announced for him. The same logic now also applies to the first policy-making stage. The principal observes no new information between the first and second policy-making stages. Hence, to avoid switching costs, at the first policy-making stage, the principal will announce whatever rule she intends to ultimately use. That rule will be the one that, when best responded to, maximizes the principal’s expected payoff.

Lemma C.1 in Appendix C gives necessary and sufficient conditions for a strategy profile to

be an equilibrium. For notational convenience we superscript equilibrium behavior by  $NC$  (for no clarity).

Suppose, in the second policy-making stage, the principal announces retention rules  $m_2^1$  (for agent 1) and  $m_2^2$  (for agent 2). Agent 2 knows that the principal will follow this rule at the retention stage. Hence, if agent 2 believes that agent 1 took the action  $\tilde{a}_1$ , then agent 2 solves

$$\max_{a_2 \in [0, \frac{\bar{a}}{2}]} \left[ (p(\tilde{a}_1, a_2)m_2^2(s) + (1 - p(\tilde{a}_1, a_2))m_2^2(f)) \underline{B} + u \left( \frac{\bar{a}}{2} - a_2 \right) \right].$$

Agent 2's best response is given by the following first-order condition:

$$p_2(\tilde{a}_1, a_2^*) (m_2^2(s) - m_2^2(f)) \underline{B} = u' \left( \frac{\bar{a}}{2} - a_2^* \right).$$

Given this, the principal will announce a retention rule  $m_2^2(s) = 1$  and  $m_2^2(f) = 0$ , since this rule maximizes  $a_2^*$  for any belief  $\tilde{a}_1$ . Call the choice induced by this retention rule  $a_2^{**}(\tilde{a}_1)$ .

Suppose, in the first policy-making stage, the principal announces the retention rules  $m_1^1$  (for agent 1) and  $m_1^2$  for agent 2. Agent 1 knows that the principal will follow the rule  $m_1^1$  at the retention stage, since she will never have an incentive to revise it in the second policy-making stage (since agent 1's actions are already taken at that point). Agent 1 also knows that, regardless of  $m_1^2$ , in the second policy-making stage, the principal will choose  $m_2^2$  as described above and (since  $a_2^{**}$  is not a function of the actual  $a_1$  chosen, which is unobservable) agent 2 will choose  $a_2^{**}(\tilde{a}_1)$ . Hence, agent 1 solves:

$$\max_{a_1 \in [0, \frac{\bar{a}}{2}]} \left[ (p(a_1, a_2^{**}(\tilde{a}_1))m_1^1(s) + (1 - p(a_1, a_2^{**}(\tilde{a}_1)))m_1^1(f)) \underline{B} + u \left( \frac{\bar{a}}{2} - a_1 \right) \right].$$

Agent 1's best response is given by the following first order condition:

$$p_1(a_1^*, a_2^{**}(\tilde{a}_1))(m_1^1(s) - m_1^1(f)) \underline{B} = u' \left( \frac{\bar{a}}{2} - a_1^* \right).$$

In the first policy-making stage, the principal will announce (and ultimately follow) a retention rule  $m_1^1(s) = 1$  and  $m_1^1(f) = 0$ , since this rule maximizes  $a_1^*$ . To avoid having to revise her rule for agent 2, the principal will also, of course, announce  $m_1^2(s) = 1$  and  $m_1^2(f) = 0$ .

Given all of this, in equilibrium, both agents are retained if the policy succeeds and neither agent is retained if the policy fails. Anticipating this, the agents choose investments to balance the marginal benefit of increased probability of success and consequently retention (given a belief

about the other agent’s action), against the marginal cost of the foregone rents from resources expended on policy. In equilibrium, beliefs are correct and equilibrium choices are characterized by the following first-order conditions (with equilibrium retention rules substituted in):

$$p_1(a_1^{NC}, a_2^{NC})\underline{B} = u' \left( \frac{\bar{a}}{2} - a_1^{NC} \right), \quad (1)$$

and

$$p_2(a_1^{NC}, a_2^{NC})\underline{B} = u' \left( \frac{\bar{a}}{2} - a_2^{NC} \right). \quad (2)$$

By the assumptions on the limits of the marginals of  $p(\cdot, \cdot)$  any solution is interior and at least one interior solution exists.

**Proposition 4.1** *Suppose there are two agents and actions are unobservable. All equilibria have the following properties on the equilibrium path:*

- (i) *The principal retains both agents with probability 1 if the policy succeeds. Otherwise the principal does not retain either agent;*
- (ii) *Agent 1 chooses  $a_1^{NC}$  consistent with equation 1 and agent 2 chooses  $a_2^{NC}$  consistent with equation 2.*

**Proof.** Follows from Lemma C.1 and the argument in the text. ■

With no clarity of responsibility—i.e., no action observability and divided agency—the principal rewards policy success by retaining both agents and punishes policy failure by replacing both agents. As such, the complete absence of clarity of responsibility introduces a “team production” problem among the agents. Both agents’ actions affect the outcome and both of their rewards depend on the outcome. Hence, the agents have incentives to free ride on each other because they do not fully internalize the positive externalities of their resource expenditures.<sup>5</sup> Given this team-production problem, equilibrium behavior by the agents when there is no clarity of responsibility

---

<sup>5</sup>To see the free rider problem formally, note first that from the standpoint of maximizing the agents’ joint welfare, their optimal choices are solutions to the problem that maximizes the sum of their individual utilities, That is,

$$\max_{(a_1, a_2)} \left[ p(a_1, a_2)\underline{B} + u_1 \left( \frac{\bar{a}}{2} - a_1 \right) + p(a_1, a_2)\underline{B} + u_2 \left( \frac{\bar{a}}{2} - a_1 \right) \right].$$

Taking the first-order conditions and gathering terms, we obtain

$$\begin{cases} p_1(a_1^*, a_2^*)2\underline{B} = u'(\bar{a}_1 - a_1^*) \\ p_2(a_1^*, a_2^*)2\underline{B} = u'(\bar{a}_2 - a_2^*). \end{cases} \quad (3)$$

leaves considerable room for improvement and points to the intuitive appeal of introducing clarity of responsibility.

## 5 Full Clarity of Responsibility

In this section, we analyze behavior under full clarity of responsibility—unified agency and action observability. We show that the retention rule the principal uses in equilibrium is characterized by WHYDFML behavior—creating incentives for the agent with potentially deleterious welfare consequences for the principal.

Before turning to the analysis, it is worth noting that, although our model’s predictions about the welfare effects of clarity of responsibility do not conform with standard intuitions, this is not because we have failed to capture the mechanisms on which those intuitions rest. As we have just shown, the complete absence of clarity of responsibility gives rise to a team production problem in our model. Moreover, as we show in Appendix B, if the two actions (i.e.,  $a_1$  and  $a_2$ ) are chosen simultaneously rather than sequentially, then clarity of responsibility has precisely the positive effects hypothesized in the literature. *It is the introduction of a sequential policy-making process that turns action observability and unified agency into negative complements.*

Consider now the game with full clarity of responsibility (one agent, observable actions) and a sequential policy-making process. A strategy for the principal is a triple,  $(m_1, m_2, \rho)$ , where  $m_1(\cdot, \cdot, \cdot) \in \mathcal{R}$  is an initial declaration of a retention rule,  $m_2(m_1, a_1)(\cdot, \cdot, \cdot) \in \mathcal{R}$  is the retention rule that will be declared in the second policy-making stage following a history  $(m_1, a_1)$ , and  $\rho(m_1, m_2, a_1, a_2, O)(\cdot, \cdot, \cdot) \in \mathcal{R}$  is the retention rule that will be used following a history  $(m_1, m_2, a_1, a_2, O)$ . A strategy for the agent is a pair  $(a_1(\cdot), a_2(\cdot, \cdot, \cdot))$  where  $a_1(m_1)$  is the action that is taken following an initial declaration  $m_1$  and  $a_2(m_1, m_2, a_1)$  is the action taken following a history  $(m_1, m_2, a_1)$ . We denote equilibrium choices with the superscript *FC* (for *full clarity*).

Lemma C.2 in Appendix C gives necessary and sufficient conditions for a strategy profile in this game to be an equilibrium. The basic logic is this. Regardless of what happens in the first policy-making process, comparing these conditions to the equilibrium conditions (1) and (2), it is clear that the choices in (3), which maximize the agents’ joint welfare, are derived by taking into account a larger marginal benefit of investing into policy, while holding fixed the marginal costs. Consequently, those choices must entail a greater policy investment by the agents.

making stage, in the second policy-making stage the principal will announce (and then follow) a retention rule that, when best responded to by the agent, maximizes the second allocation. Anticipating this fact, in the first policy-making stage the agent will only find an announced rule credible if the principal will have no incentive to revise the rule in the second policy-making stage. Hence, in the first policy-making stage, the principal will announce a rule that (when best responded to by the agent) maximizes the first allocation, subject to the constraint that it then maximizes the second allocation at all histories. Given this, we solve for equilibrium behavior starting at the end of the game.

Before turning to the analysis, one further piece of notation will be useful.

**Definition 5.1** *For a given total level of spending,  $A \leq \bar{a}$ , define the efficient division as follows:*

$$(a_1^{\text{efficient}}(A), a_2^{\text{efficient}}(A)) = \arg \max_{(a_1, a_2)} p(a_1, a_2) \text{ subject to } a_1 + a_2 = A.$$

The most powerful incentives the principal can create to choose some  $a'_2$  is to offer certain retention for any  $a_2 = a'_2$  and certain non-retention for any  $a_2 \neq a'_2$ . Hence, at a history  $(m_1, a_1)$ , the principal can induce the agent to choose  $a'_2$  if and only if:

$$\bar{B} + u(\bar{a} - a_1 - a'_2) \geq u(\bar{a} - a_1).$$

This implies that at a history  $(m_1, a_1)$  the principal can induce any  $a_2$  less than or equal to  $\hat{a}_2(a_1)$  given by:

$$\bar{B} + u(\bar{a} - a_1 - \hat{a}_2(a_1)) = u(\bar{a} - a_1). \quad (4)$$

Given this, at a history  $(m_1, a_1)$ , there is a rule that induces the agent to spend the full remaining budget (i.e., choose  $a_2 = \bar{a} - a_1$ ) if and only if  $a_1 \geq \hat{a}_1$  implicitly defined by:

$$\bar{B} + u(0) = u(\bar{a} - \hat{a}_1).$$

This means the principal can extract the entire budget of  $\bar{a}$  if  $\bar{B} + u(0) \geq u(\bar{a})$ . Moreover, if he is spending the entire budget, the agent is indifferent as to its division between  $a_1$  and  $a_2$ . As shown in the next result, in this eventuality, the principal will adopt a rule that induces the agent to choose  $(a_1^{\text{efficient}}(\bar{a}), a_2^{\text{efficient}}(\bar{a}))$ .

**Lemma 5.1** *Consider the game with full clarity of responsibility. Suppose  $\bar{B} + u(0) \geq u(\bar{a})$ . Then there exist a retention rule consistent with equilibrium such that the agent's best response is to choose  $(a_1^{\text{efficient}}(\bar{a}), a_2^{\text{efficient}}(\bar{a}))$ .*

**Proof.** See Appendix A. ■

Now consider the case where  $\bar{B} + u(0) < u(\bar{a})$ . Here we must think about two cases: histories with  $a_1 \geq \hat{a}_1$  and histories with  $a_1 < \hat{a}_1$

Consider a history with  $a_1 < \hat{a}_1$ , so that no rule the principal can declare in the second policy-making period will induce the agent to spend the whole remaining budget (i.e., choose  $a_2 = \bar{a} - a_1$ ). At such a history, the most the principal can extract at the second policy-making stage is  $\hat{a}_2(a_1)$ . She does so by announcing that she will retain for certain if  $a_2 = \hat{a}_2(a_1)$  and not retain if  $a_2 < \hat{a}_2(a_1)$ .<sup>6</sup>

Now consider a history with  $a_1 \geq \hat{a}_1$ , so that the principal can induce the agent to spend the whole remaining budget (i.e., choose  $a_2 = \bar{a} - a_1$ ). Clearly, following such a history, the principal will announce a rule that does in fact induce the agent to spend the whole remaining budget.

Given this, consider the agent's incentives in the first policy-making stage. If the agent chooses  $a_1 < \hat{a}_1$ , he knows the principal will announce a rule in the second period that induces  $\hat{a}_2(a_1)$  and that retains him for certain for doing so. Hence, the agent's payoff will be:

$$\bar{B} + u(\bar{a} - a_1 - \hat{a}_2(a_1)).$$

Given this, if the agent is going to choose  $a_1 < \hat{a}_1$ , he wants to minimize  $a_1 + \hat{a}_2(a_1)$ . The next result shows that the agent does this by choosing  $a_1 = 0$ .

**Lemma 5.2** *Consider the game with full clarity of responsibility. For any  $m_1$  consistent with equilibrium,  $a_1 = 0$  dominates any other  $a_1 < \hat{a}_1$  for the agent.*

**Proof.** See Appendix A. ■

The agent's expected payoff from choosing  $a_1 = 0$  is:

$$\bar{B} + u(\bar{a} - \hat{a}_2(0)).$$

If, instead, the agent chooses  $a_1 \geq \hat{a}_1$ , the principal will induce the agent to choose  $a_2 = \bar{a} - a_1$  in the second policy-making stage. Hence, the agent's payoff from choosing  $a_1 \geq \hat{a}_1$  is bounded

---

<sup>6</sup>She can announce whatever she likes for choices  $a_2 > \hat{a}_2(a_1)$ , the agent will never choose such an allocation.

above by  $\bar{B} + u(0)$ , which is less than  $\bar{B} + u(\bar{a} - \hat{a}_2(0))$ , the payoff from  $(0, \hat{a}_2(0))$ . This implies that, when  $\bar{B} + u(0) < u(\bar{a})$ , the equilibrium allocations are  $(0, \hat{a}_2(0))$ .

**Proposition 5.1** *Consider the game with full clarity of responsibility. In any SPNE that is robust to small revision costs, on the equilibrium path, allocations are:*

$$\begin{cases} (0, \hat{a}_2(0)) & \text{if } \bar{B} < u(\bar{a}) - u(0) \\ (a_1^{\text{efficient}}(\bar{a}), a_2^{\text{efficient}}(\bar{a})) & \text{if } \bar{B} > u(\bar{a}) - u(0). \end{cases}$$

*In the knife-edge case where  $\bar{B} = u(\bar{a}) - u(0)$ , any allocation  $(a_1, \bar{a} - a_1)$  can occur in equilibrium.*

**Proof.** When  $\bar{B} \neq u(\bar{a}) - u(0)$ , the argument follows from the argument in the text. In the knife-edge case, as shown in the proof of Lemma 5.1, any division of the full budget is a best response from the agent. ■

The intuition behind the equilibrium behavior described in Proposition 5.1 is as follows. The principal must always extract as much investment at the second allocation as possible with her retention rule. In the first policy-making stage, were she to announce a retention rule that would not do so, she would end up wanting to revise it. Hence, only if the principal is able to extract the entire budget at the second allocation can she credibly create incentives for the first allocation.

As Proposition 5.1 shows, the agent's division of resources under full clarity of responsibility need not be efficient with respect to maximizing the probability of success, given a level of expenditures. In particular, unless the agent is spending the full budget, all the attention is on the second allocation. Put differently, the equilibrium described in Proposition 5.1 has what Weingast, Shepsle and Johnsen (1981) refer to as the *what have you done for me lately* (WHYDFML) property. This over-emphasis on the late stage of the policy-making process holds under clarity of responsibility even if the first allocation has a larger marginal impact on the probability of success than does the second allocation. When policy making is sequential, clarity of responsibility can, thus, create incentives with negative welfare consequences for the principal.

Indeed, putting some additional structure on the functional forms of  $p$  and  $u$ , we can give precise conditions under which this is the case.

**Proposition 5.2** *Let  $u(x) = x$  and  $p(a_1, a_2) = \gamma\sqrt{a_1} + (1 - \gamma)\sqrt{a_2}$ , with  $\gamma \in (0, 1)$  and  $\bar{a} = 1$ . Then the following are true:*

- (i) *If  $\bar{B} < \bar{a}$ , then there exists a  $\gamma' < 1$ , such that no clarity of responsibility is preferred to full clarity of responsibility if and only if  $\gamma \geq \gamma'$ .*
- (ii) *If  $\bar{B} \geq \bar{a}$ , then full clarity of responsibility is preferred to no clarity of responsibility for all  $\gamma \in (0, 1)$ .*

**Proof.** See Appendix A. ■

## 6 Eliminating Clarity

Intuitively, WHYDFML seems to be driven precisely by the conjunction of the two institutional features that characterize clarity of responsibility. Action observability means that the principal knows “too much” about the policy process, which prevents her from punishing departures from good behavior in the first allocation. The presence of a unitary agent prevents the principal from providing separate incentives for each allocation. In what follows we show that, at least in some situations, eliminating either one of these features of clarity of responsibility can make the principal better off by eliminating WHYDFML behavior. To do so, we first consider the implications for principal welfare of removing these two institutional features one at a time, and then compare them to the implications of eliminating clarity of responsibility entirely.

### 6.1 Eliminating Action Observability

Consider the game in which there is still a single agent, but the principal observes just the outcome and not the agent’s actions. We superscript the equilibrium choices in this game by *NO* for *no observability*.

A strategy for the principal is a triple,  $(m_1, m_2, \rho)$ , where  $m_1(\cdot, \cdot, \cdot) \in \mathcal{R}$  is an initial declaration of a retention rule,  $m_2(m_1, a_1)(\cdot, \cdot, \cdot) \in \mathcal{R}$  is the retention rule that will be declared in the second policy-making stage following a history  $(m_1, a_1)$ , and  $\rho(m_1, m_2, a_1, a_2, O)(\cdot, \cdot, \cdot) \in \mathcal{R}$  is the retention rule that will be used following a history  $(m_1, m_2, a_1, a_2, O)$ . Given that actions are not observable, the principal’s strategy must be constant in  $a_1$  and  $a_2$ . In light of this, we write retention rules only

as mappings from outcomes into retention probabilities, suppressing the dependency on actions. Moreover, we will write the choice of  $m_2$  only as a function of the observed  $m_1$  and the choice of  $\rho$  only as a function of the observed  $m_1, m_2$ , and the outcome (since the rest of the histories are unobserved).

A strategy for the agent is a pair  $(a_1, a_2)$  where  $a_1(m_1)$  is the action that is taken following an initial declaration  $m_1$  and  $a_2(m_1, m_2, a_1)$  is the action taken following a history  $(m_1, m_2, a_1)$ .

Lemma C.3 in Appendix C gives necessary and sufficient conditions for a strategy profile in this game to be an equilibrium. Regardless of what happens in the first policy-making stage, in the second policy-making stage the principal will announce (and then follow) a retention rule that, when best responded to, maximizes the second allocation. Anticipating this fact, in the first policy-making stage the agent will only find an announced rule credible if the principal will have no incentive to revise the rule. Hence, in the first policy-making stage, the principal will announce a rule that (when best responded to) maximizes the first allocation, subject to the constraint that it then maximizes the second allocation at all histories. Unlike in the game with full clarity of responsibility, the principal cannot condition her response on the second stage action here. As such, a rule that maximizes the principal's expected welfare at the first policy-making stage will continue to do so in the second policy-making stage.

Given this, equilibrium behavior is straightforward to characterize in this game. We can represent a retention rule with two values: the probability of retention given a successful outcome (labeled  $r(s)$ ) and the probability of retention given a failed outcome (labeled  $r(f)$ ). We solve the game starting at the end.

If the principal announces  $m_2$ , the agent's problem in the second policy-making stage is:

$$\max_{a_2} (p(a_1, a_2)m_2(s) + (1 - p(a_1, a_2))m_2(f))\bar{B} + u(\bar{a} - a_1 - a_2).$$

The agent's best response is given by the following first-order condition:

$$p_2(a_1, a_2^*)(m_2(s) - m_2(f))\bar{B} = u'(\bar{a} - a_1 - a_2^*). \quad (5)$$

In the second policy-making stage, the agent balances the marginal benefit, in terms of retention probability, of expenditures on public policy against the marginal costs, in terms of forgone rents.

It is clear from Equation 5 that  $a_2^*$  is increasing in  $m_2(s)$  and decreasing in  $m_2(f)$ , regardless of  $a_1$ . Hence, in any equilibrium,  $m_2(m_1)(s) = 1$  and  $m_2(m_1)(f) = 0$  for all  $m_1$ . This implies that

the agent's second allocation is characterized by:

$$p_2(a_1, a_2^*)\bar{B} = u'(\bar{a} - a_1 - a_2^*).$$

Assumptions on  $p$  guarantee that the optimum exists and is unique.

The analysis above shows that the  $m_2$  the principal will announce is unresponsive to the previous announcement  $m_1$ . Consequently, in the first policy-making stage, the agent knows that ultimately he will be retained if policy succeeds and replaced if policy fails. Given this, for all  $m_1$ , the agent's best response in the first policy-making stage is given by:

$$p_1(a_1^*, a_2^*(a_1^*))\bar{B} = u'(\bar{a} - a_1^* - a_2^*(a_1^*)),$$

And, since the announcement  $m_1$  has no effect on behavior, the principal will announce  $m_1$  identical to  $m_2$ , to avoid the costs of revision.

Notice, in equilibrium, given a total level of expenditure ( $a_1^{NO} + a_2^{NO}$ ) the agent spends the resources efficiently—i.e.,  $p_1(a_1^{NO}, a_2^{NO}) = p_2(a_1^{NO}, a_2^{NO})$ . This is because, conditional on the amount allocated, the agent wants to maximize the probability of a successful policy outcome so as to maximize the probability of retention.

The discussion above leads to the following characterization of equilibrium play.

**Proposition 6.1** *Consider the game with unobservable actions and a single agent. Any equilibrium has the following properties on the equilibrium path:*

- (i) *The principal retains the agent if and only if policy succeeds.*
- (ii) *The agent chooses allocations implicitly defined by:*

$$p_1(a_1^{NO}, a_2^{NO})\bar{B} = p_2(a_1^{NO}, a_2^{NO})\bar{B} = u'(\bar{a} - a_1^{NO} - a_2^{NO}). \quad (6)$$

**Proof.** Follows from Lemma C.3 and the argument in the text. ■

### Comparison to Full Clarity of Responsibility

When policy making is sequential, the welfare effects of eliminating action observability are subtle. The institutional features that give rise to WHYDFML behavior under full clarity of responsibility can be bad for the principal because they lead to an inefficient distribution of resources that focuses

too much on the late stage of the policy-making process. Proposition 6.1 shows that eliminating action observability restores efficiency by eliminating WHYDFML.

This increased efficiency, however, comes at cost—the total resources expended on policy are lower without action observability. This is because the principal cannot offer as powerful incentives when she can only condition her retention rule on noisy outcomes, rather than on actual actions.

The following remark formalizes the result.

**Remark 6.1** *In equilibrium, total agent expenditures are higher in the game with full clarity of responsibility than in the game with unified agency but no action observability.*

**Proof.** See Appendix A. ■

The net effect of eliminating action observability, thus, depends on the trade-off between increased efficiency (due to eliminating WHYDFML) and decreased total expenditures. Again, putting some additional structure on the functional forms of  $p$  and  $u$ , we can characterize which effect dominates when.

**Proposition 6.2** *Suppose that  $u(x) = x$  and  $p(a_1, a_2) = \gamma\sqrt{a_1} + (1 - \gamma)\sqrt{a_2}$ , with  $\gamma \in (0, 1)$  and  $\bar{a} = 1$ . Then the following are true:*

- (i) *If  $\bar{B} < \bar{a}$ , then there exists a  $\gamma^* < 1$ , such that unified agency with no action observability is preferred to full clarity of responsibility if and only if  $\gamma \geq \gamma^*$ .*
- (ii) *If  $\bar{B} \geq \bar{a}$ , then full clarity of responsibility is preferred to unified agency with no action observability for all  $\gamma \in (0, 1)$ .*

**Proof.** See Appendix A. ■

The first point of Proposition 6.2 correspond to the intuition above regarding the inefficiencies associated with WHYDFML. If  $\gamma$  is large, the first allocation is particularly important, so WHYDFML is particularly damaging. In this case, as long as the incentive-power effect is not too large, eliminating action observability is good for the principal. The second point shows that, when the benefits of holding office are large enough, the stronger incentives associated with full clarity of responsibility imply that WHYDFML concerns cease to matter because the agent is able to extract the full budget.

Proposition 6.2 has two implications that may be counterintuitive given previous work on clarity of responsibility. First, when policy making is sequential, the principal may be better off relaxing clarity of responsibility by eliminating action observability. Second, the principal may be better off if she is able to forego observing the agent’s choices at a time when she is *most* concerned about how the agent divides the expenditures across allocations (i.e., when  $\gamma$  is close to one). This is because, in the absence of action observability, the agent will choose a division of resources efficiently. Hence, when the marginal returns to one action are much larger than another, without action observability this asymmetry will be reflected in the agent’s division of the resources between the two actions.

## 6.2 Eliminating Unified Agency

In this section, we consider the game in which there are two agents and allocation decisions are observable to all players. The principal now makes separate retention decisions regarding each of the two agents. Hence, in a policy making stage  $t$ , the principal declares a separate retention rule for each agent and can, conceivably, revise zero, one, or both rules in each subsequent stage. For convenience, we will write  $m_t = (m_t^1, m_t^2)$ , where  $m_t^i$  is the declared retention rule for agent  $i$ . Similarly, at the retention stage we will write  $\rho = (\rho^1, \rho^2)$ . We superscript the equilibrium values for this game with *NU* for *no unified agency*.

Lemma C.4 in Appendix C gives necessary and sufficient conditions for a strategy profile in this game to be an equilibrium. Regardless of what happens in the first policy-making stage, in the second policy-making stage the principal will declare (and follow) a retention rule for the second agent that (when best responded to) maximizes the second agent’s allocation. The principal will never revise the retention rule she announced for the first agent, since his actions are already taken. Given this, in the first policy-making stage, the principal will announce (and later follow) a retention rule for the first agent that (when best responded to) maximizes the first agent’s allocation. To avoid revision costs, in the first policy-making stage, the principal will of course announce the same rule for the second agent that she intends to announce for the second agent in the second policy-making stage.

Given this, we can simply study the principal’s retention rules for the two agents separately. For each agent, the principal must identify the rule that maximizes that agent’s allocation and announce this in both policy-making stages. To achieve this, the principal will use retention rules

that condition only on a given agent's allocation. Conditioning on the other agent's action or on outcomes can only weaken incentives by adding team production and noise.

Agent  $i$ 's payoff from allocating  $a_i$  and being retained with certainty is  $\underline{B} + u(\frac{\bar{a}}{2} - a_i)$ . Her payoff from allocating nothing and not being retained is  $u(\frac{\bar{a}}{2})$ . Define  $\check{a}$  as the level of spending that satisfies:

$$\underline{B} + u\left(\frac{\bar{a}}{2} - \check{a}\right) = u\left(\frac{\bar{a}}{2}\right). \quad (7)$$

The principal can induce agent  $i$  to allocate any amount  $a'_i \leq \min\{\frac{\bar{a}}{2}, \check{a}\}$  by threatening not to retain if that agent does not allocate precisely  $a'_i$ . In equilibrium, therefore, the principal will use a retention rule that induces each agent  $i$  to allocate precisely  $\min\{\frac{\bar{a}}{2}, \check{a}\}$ . This will lead to full resource extraction (i.e.,  $a_i^{NU} = \frac{\bar{a}}{2}$ ) when  $\underline{B} + u(0) \geq u(\frac{\bar{a}}{2})$ . We then have the following result:

**Proposition 6.3** *Consider the game with observable actions and two agents. Any equilibrium has the following properties on the equilibrium path:*

- (i) *If  $\underline{B} + u(0) < u(\frac{\bar{a}}{2})$ , each agent chooses  $a_i^{NU} = \check{a}_i$ , implicitly defined in equation 7;*
- (ii) *If  $\underline{B} + u(0) \geq u(\frac{\bar{a}}{2})$ , each agent chooses  $a_i^{NU} = \frac{\bar{a}}{2}$ .*

**Proof.** Follows from Lemma C.4 and the argument in the text. ■

### Comparison to Full Clarity of Responsibility

The equilibrium above suggests that moving from a unitary agent to multiple agents has two effects on the principal's welfare—one positive and one negative.

The positive effect has to do with eliminating the WHYDFML behavior that occurs with full clarity of responsibility. When there are two agents, the principal can give them separate incentives. As such, eliminating unified agency removes the commitment problem that leads the principal to use a retention rule that induces the agent to over emphasize the late stage of the policy-making process under full clarity of responsibility.

The negative effect comes from the agents' increased incentives for rent seeking. This increase in rent seeking has two sources. First, holding office is worth less to an agent when authority is divided. Second, agents control fewer resources under divided authority. Since, agents have diminishing marginal utility from rents, controlling fewer resources makes rent seeking more attractive.

Whether eliminating unified agency benefits the principal depends on the magnitudes of these competing effects. Restricting  $p$  to be additively separable in  $a_1$  and  $a_2$ , we can give precise conditions under which eliminating unified agency does and does not benefit the principal.

**Proposition 6.4** *Suppose that  $p(a_1, a_2) = \gamma f(a_1) + (1 - \gamma)g(a_2)$ , with  $\gamma \in (0, 1)$  and  $f, g$  increasing, concave, mapping into  $(0, 1)$ , and satisfying the Inada conditions. Then, in equilibrium, the principal's welfare is higher with divided agency than with full clarity of responsibility if and only if both of the following hold:*

(i)  $\bar{B} \leq u(\bar{a}) - u(0)$  and

(ii)  $\gamma$  sufficiently large.

**Proof.** See Appendix A. ■

The intuitions for this result invoke several facts about actors' incentives noted above. When the benefits of holding office under full clarity of responsibility are sufficiently large ( $\bar{B} \geq u(\bar{a}) - u(0)$ ), the principal can extract the full budget and induce an efficient division. As such, in this case, full clarity of responsibility is an optimal institution.

However, when the benefits of holding office under full clarity of responsibility do not allow the principal to extract the full budget, the principal faces a trade-off. On the one hand, dividing responsibility among two agents eliminates WHYDFML behavior. On the other hand, incentives are weaker under divided agency. As a result, the principal is better off under full clarity of responsibility if the first allocation is not too important, so that the inefficiencies associated with WHYDFML do not loom too large. And the principal is better off under divided agency if the first allocation is important.

### 6.3 No Clarity versus Partial Clarity

Recall that, as shown in Proposition 5.2, if the WHYDFML effect looms large, then it is possible for the principal to prefer having no clarity of responsibility to having fully clarity of responsibility. This raises the question of whether no clarity of responsibility could potentially be the optimal institution. Here we show that this cannot be the case. The principal always prefers some limited form of clarity of responsibility to no clarity of responsibility at all.

To see this, compare the game with no clarity of responsibility to the game with observable actions but divided agency. In the game with two agents and observable actions there is no WHYDFML and the principal is able to provide more powerful incentives than she is with no clarity of responsibility, since she can condition her retention rule on actual actions, not just outcomes. Thus, while full clarity of responsibility is not always the optimal institution (see Propositions 6.2 and 6.4), no clarity of responsibility is never the optimal institution, as shown in the next result.

**Proposition 6.5** *Eliminating both aspects of clarity of responsibility is never the optimal institutional arrangement for the principal.*

**Proof.** See Appendix A. ■

## 7 Discussion

Clarity of responsibility has emerged as one of the key institutional factors in explaining patterns of economic and political outcomes in comparative politics. Earlier work has focused on the positive impacts of clarity of responsibility. In contrast to that prior work, our analysis suggests that, with sequential policy making, clarity of responsibility comes with trade-offs. Complete clarity of responsibility increases the costly actions agents take to achieve good policy outcomes, but at the same time it creates WHYDFML behavior. As such, whether clarity of responsibility promotes the principal's welfare is contingent on other features of the policy-making environment, for instance, the relative importance of different stages of the policy-making process. In what follows, we relate the implications of our models to several issues that have been the focus of considerable attention from social scientists and policy analysts.

### 7.1 Fire Alarms vs. Police Patrols

As McCubbins and Schwartz (1984) note, legislative oversight can be thought of as corresponding to two distinct informational regimes: *police patrols*, in which the overseer actively supervises the agent's actions and *fire alarms*, in which overseer involvement is prompted by bad policy outcomes. McCubbins and Schwartz argue that the fire alarm regime may be more efficient: interest groups

and other concerned third parties have incentives to pay close attention to agency choices and bring them to Congress’s attention when the outcomes are particularly damaging.

Our analysis suggests a complementary rationale for fire alarms. A police patrol creates clarity of responsibility, while a fire alarm eliminates action observability. Our analysis, thus, suggests that the active oversight associated with police patrols can give rise to WHYDFML behavior—inducing agencies to over-emphasize the later stages of the policy-making process at the expense of the earlier stages. Moving to a fire alarm regime can, in contrast, eliminate the inefficiencies induced by WHYDFML behavior, though at the cost of decreased total expenditures or effort devoted to policy and increased rent seeking by the agency.

## 7.2 Design of Regulatory Agencies

Regulatory agencies are often tasked with both creating the regulatory regime (e.g., specifying the standards, identifying the range of remedies,) and implementing or enforcing that regime. Our model suggests that the institutional arrangement that maximizes accountability may vary across different agencies, depending on the relative importance of the early versus late stages of the policy-making process.

Two U.S. agencies provide contrasting examples. The SEC is charged with the regulation of the financial sector, including defining what constitutes financial improprieties. While there may be disagreements about stricter versus laxer regulations, defining acceptable financial practices is considerably less difficult than detecting transgressions (e.g., insider trading, abusive short-selling practices, investment fraud). Hence, WHYDFML behavior that induces the SEC bureaucracy to overemphasize enforcement may not be terribly costly. In contrast, the technologies used by state and federal EPAs to monitor compliance with Clean Water Act regulations have much higher reliability and are generally easy to apply, while codifying technology-based standards (e.g., effluent guidelines, categorical pretreatment and secondary treatment standards) and water quality standards is complex and contentious. Hence, WHYDFML behavior that induces the EPA bureaucracy to underemphasize the drafting of regulations may be highly costly. Given this, our results suggest that the benefits (in terms of decreased rent seeking) of clarity of responsibility may outweigh the costs (in terms of WHYDFML) for agencies, like the SEC, where enforcement is paramount, but not for agencies like the EPA, where the drafting of high quality regulations is paramount.

### 7.3 Design of Policy

When the institutional environment is fixed, our model suggests a second-best approach to policy design. If faced with an institution that creates incentives for WHYDFML behavior, policy makers should look for policy solutions where the inefficiencies created by such behavior are not particularly costly.

For instance, suppose an agency designing environmental regulations can take one of two approaches. It can set emissions-targets and engage in costly monitoring or it can impose a technology requirement on firms. The former policy requires effort late in the process, in the form of monitoring. The latter policy requires effort early in the process, in the form of choosing the right technology requirements.

Suppose that the first-best policy involves technology requirements, not emissions-targets and monitoring. Nonetheless, if the agency that will implement the policy is characterized by full clarity of responsibility, a legislature might want to instruct the agency to follow a policy of emissions-targets and monitoring as a second-best response to the incentives associated with WHYDFML. The reason is that, following the logic of our model, the legislature can anticipate that the agency will engage in inefficient behavior—focusing too much on the later stages of the policy-making process. As such, it wants to select a policy that will not be too severely negatively affected by that behavior, even if such policies are not the first best.<sup>7</sup>

## 8 Conclusion

In a sequential policy-making environment, full clarity of responsibility gives rise to WHYDFML behavior, which induces the agent to over-emphasize the later stages of the policy-making process relative to the early stages. Eliminating either of the two institutional features that comprise clarity of responsibility—action observability or unified agency—eliminates WHYDFML behavior and the inefficiencies it induces, but at the cost of increased rent seeking. As such, whether complete clarity of responsibility, or some more limited form of clarity of responsibility, is optimal depends on the

---

<sup>7</sup>Other areas of policy debate for which our results may be relevant but which, given space constraints, we leave here without detailed discussion, include the desirability of strict disclosure requirements for financial statements by management of public corporations, which became law in the U.S. under Sarbanes-Oxley Act of 2002.

specifics of the environment and, in particular, the relative importance of the various stages of the policy-making process. Eliminating both aspects of clarity of responsibility is never optimal, since doing so introduces an additional team production problem among the agents. However, it is possible for the principal to prefer no clarity of responsibility to complete clarity of responsibility.

Taken together, our results suggest that the setting with sequential policy making presents a striking mix of incentives for both principals and agents that significantly complicates both normative institutional prescriptions and empirical analyses of the effects of clarity of responsibility. It is our hope that our results on the differential impacts of the two institutional components of clarity of responsibility will stimulate empirical work that attempts to disentangle and quantify the competing effects in settings with sequential policy making.

## Appendix A Proofs of Numbered Results

The following gives the formal definition of robustness to small revision costs.

**Definition A.1** *Consider a game with two policy-making stages. An associated SPNE,  $s^* = (m_1^*, m_2^*(m_1, a_1), \rho^*(m_1, m_2, a_1, a_2, O), a_1^*(m_1), a_2^*(m_1, m_2, a_1))$ , is **robust to small revision costs** if, for all  $\delta > 0$ , there exists an  $\bar{\epsilon} > 0$  such that, for all  $\epsilon < \bar{\epsilon}$ , there is an equilibrium,  $s^\epsilon = (m_1^\epsilon, m_2^\epsilon(m_1, a_1), \rho^\epsilon(m_1, m_2, a_1, a_2, O), a_1^\epsilon(m_1), a_2^\epsilon(m_1, m_2, a_1))$ , of the perturbed game, satisfying:*

(i)  $|m_1^\epsilon(a_1, a_2, O) - m_1^*(a_1, a_2, O)| < \delta$  for all  $(a_1, a_2, O)$ ;

(ii)  $|m_2^\epsilon(m_1, a_1)(a_1, a_2, O) - m_2^*(m_1, a_1)(a_1, a_2, O)| < \delta$  for all  $(m_1, a_1)$  and  $(a_1, a_2, O)$ ;

(iii)  $|\rho^\epsilon(m_1, m_2, a_1, a_2, O)(a_1, a_2, O) - \rho^*(m_1, m_2, a_1, a_2, O)(a_1, a_2, O)| < \delta$  for all  $(m_1, m_2, a_1, a_2, O)$ ;

(iv)  $|a_1^\epsilon(m_1) - a_1^*(m_1)| < \delta$  for all  $m_1$ ; and

(v)  $|a_2^\epsilon(m_1, m_2, a_1) - a_2^*(m_1, m_2, a_1)| < \delta$  for all  $(m_1, m_2, a_1)$ .

### Proof of Lemma 5.1

Consider the following retention rule:

- Certain retention for  $(a_1^{\text{efficient}}(\bar{a}), a_2^{\text{efficient}}(\bar{a}))$

- For any pair  $(a_1, \bar{a} - a_1)$ , with  $a_1 \neq a_1^{\text{efficient}}(\bar{a})$ , retain with probability  $\pi(a_1)$  satisfying:

$$\pi(a_1)\bar{B} = u(\bar{a} - a_1) - u(0).$$

- For any pair  $(a_1, a_2)$  such that  $a_1 + a_2 < \bar{a}$  do not retain.

First suppose  $\bar{B} + u(0) > u(\bar{a})$ . For any  $a_1$ , the payoff from choosing  $a_2 = \bar{a} - a_1$  is  $\bar{B} + u(0)$  and the payoff from choosing  $a_2 < \bar{a} - a_1$  is  $u(\bar{a} - a_1)$ . Since,  $\bar{B} + u(0) > u(\bar{a})$ , it is a unique best response for the agent to choose  $a_2 = \bar{a} - a_1$  for any  $a_1$ . Hence, at no history does the principal have an incentive to revise.

Now consider the first policy-making stage incentives for the agent created by this rule. The payoff from choosing  $(a_1^{\text{efficient}}(\bar{a}), a_2^{\text{efficient}}(\bar{a}))$  is  $\bar{B} + u(0)$ . The payoff from choosing any other pair  $(a_1, \bar{a} - a_1)$  is  $\pi(a_1)\bar{B} + u(0)$ . Since  $\bar{B} > u(\bar{a}) - u(0) > u(\bar{a} - a_1) - u(0)$ , clearly  $\pi(a_1) < 1$ . Hence, the agent prefers  $(a_1^{\text{efficient}}(\bar{a}), a_2^{\text{efficient}}(\bar{a}))$  over any other allocation  $(a_1, \bar{a} - a_1)$  and, as we've already seen, the agent prefers any  $(a_1, \bar{a} - a_1)$  to any allocation  $(a_1, a_2)$  that doesn't expend the whole budget.

Since the principal has at least one retention rule that achieves her first best and which she has no incentive to revise in the second policy-making period, she will clearly announce a rule that in fact achieves her first best.

Now consider the case with  $\bar{B} + u(0) = u(\bar{a})$ . In this case all arguments are the same, except that the agent is indifferent, so clearly behaving as above is a best response, though not unique. ■

## Proof of Lemma 5.2

For any history  $(m_1, a_1)$  with  $a_1 < \hat{a}_1$ , the agent anticipates an equilibrium payoff of

$$\bar{B} + u(\bar{a} - a_1 - \hat{a}_2(a_1)).$$

Differentiating with respect to  $a_1$ , the agent will choose  $a_1 = 0$  if:

$$-u'(\bar{a} - a_1 - \hat{a}_2(a_1)) \left( 1 + \frac{d\hat{a}_2(a_1)}{da_1} \right) < 0.$$

Since  $-u' < 0$ , the condition holds if  $1 + \frac{d\hat{a}_2(a_1)}{da_1} > 0$ . Thus, all that remains is to show that  $\frac{d\hat{a}_2(a_1)}{da_1} > -1$ .

Recall that  $\hat{a}_2(a_1)$  is implicitly defined by equation 4. Differentiating the implicit function we have:

$$\frac{d\hat{a}_2(a_1)}{da_1} = \frac{u'(\bar{a} - a_1) - u'(\bar{a} - a_1 - \hat{a}_2)}{u'(\bar{a} - a_1 - \hat{a}_2)}.$$

Since  $u$  is concave, this is clearly negative. However, it is greater than  $-1$  since  $u'(\bar{a} - a_1) > 0$ . ■

### Proof of Proposition 5.2

First suppose  $\bar{B} \geq \bar{a}$ . By Proposition 5.1, under full clarity the whole budget is allocated and the allocation is efficient. This is strictly better for the principal than any outcome under no clarity of responsibility, since the allocations under no clarity of responsibility are always interior.

Now suppose  $\bar{B} < \bar{a}$ . According to Proposition 5.1, under full clarity of responsibility, the first allocation is 0 and the second allocation is  $\hat{a}_2(0) = \bar{B}$ . Thus, the principal's welfare in this case is  $(1 - \gamma)\sqrt{\bar{B}}$ .

Next, consider the case with no clarity of responsibility. Taking first order conditions and rearranging shows that the first allocation is  $a_1^{NC} = \left(\frac{\gamma\bar{B}}{2}\right)^2$ , and the second allocation is  $a_2^{NC} = \left(\frac{(1-\gamma)\bar{B}}{2}\right)^2$ . Thus, the principal's welfare in this case is  $\frac{(\gamma^2 + (1-\gamma)^2)\bar{B}}{2}$ .

Comparing the two, we find that the principal's welfare is higher with no clarity of responsibility than with full clarity of responsibility if

$$\frac{\gamma^2 + (1 - \gamma)^2}{1 - \gamma} \geq \frac{2\sqrt{\bar{B}}}{\bar{B}}.$$

Since the right-hand side is greater than 1, this inequality clearly does not hold at  $\gamma = 0$  and it clearly does hold as  $\gamma$  approaches 1. Hence, it suffices to show that  $\frac{(\gamma^2 + (1-\gamma)^2)}{1-\gamma}$  is strictly convex, which is straightforward from differentiating twice. ■

### Proof of Remark 6.1

First consider the case where  $\bar{B} + u(0) < u(\bar{a})$ . In this case, under complete clarity total expenditures are  $0 + \hat{a}_2(0)$ . Under unified agency but no action observability, total expenditures are  $a_1^{NO} + a_2^{NO}$ . To see that  $\hat{a}_2(0) > a_1^{NO} + a_2^{NO}$ , note that, from the interiority of  $(a_1^{NO}, a_2^{NO})$ , we have

$$p(a_1^{NO}, a_2^{NO})\bar{B} + u(\bar{a} - a_1^{NO} - a_2^{NO}) \geq u(\bar{a}).$$

Combining this with the definition of  $\hat{a}_2(0)$  from equation 4, we have:

$$\begin{aligned} u(\bar{a} - \hat{a}_2(0)) &= u(\bar{a}) - \bar{B} \\ &< u(\bar{a}) - p(a_1^{NO}, a_2^{NO})\bar{B} \\ &\leq u(\bar{a} - a_1^{NO} - a_2^{NO}). \end{aligned}$$

Now consider the case where  $\bar{B} + u(0) \geq u(\bar{a})$ . In this case, the full budget is allocated under full clarity of responsibility, but not under unified agency but no action observability. ■

## Proof of Proposition 6.2

Taking first-order conditions and rearranging, it follows directly from Proposition 6.1 that  $a_1^{NO} = \left(\frac{\gamma\bar{B}}{2}\right)^2$  and  $a_2^{NO} = \left(\frac{(1-\gamma)\bar{B}}{2}\right)^2$ . This implies that the principal's expected payoff with unified agency and no action observability is

$$\frac{(\gamma^2 + (1-\gamma)^2)\bar{B}}{2}.$$

Now, consider the case with  $\bar{B} < \bar{a}$ . In this case, from Proposition 5.1, we have  $a_1^{FC} = 0$  and  $a_2^{FC} = \bar{B}$ . Hence the principal's expected payoff under full clarity of responsibility is:

$$(1-\gamma)\sqrt{\bar{B}}.$$

Comparing, the principal is better off without action observability if:

$$\frac{(\gamma^2 + (1-\gamma)^2)\bar{B}}{2} \geq (1-\gamma)\sqrt{\bar{B}},$$

which is equivalent to:

$$\frac{(\gamma^2 + (1-\gamma)^2)}{2(1-\gamma)} \geq \frac{1}{\sqrt{\bar{B}}}.$$

Since  $\bar{B} < \bar{a} = 1$ , the right-hand side is greater than 1. At  $\gamma = 0$  the left-hand side is equal to  $\frac{1}{2}$  and as  $\gamma$  goes to 1, the left-hand side goes to infinity. Thus, it suffices, to establish point 1, to show that the left-hand side is strictly convex. Differentiating twice and rearranging shows that this is straightforwardly true.

Now consider the case with  $\bar{B} > \bar{a}$ . Here, by Proposition 5.1, under full clarity, the principal extracts the full budget distributed efficiently. This is not the case without action observability. Thus, full clarity is preferred in this case. ■

## Proof of Proposition 6.4

By Proposition 5.1, if  $\bar{B} \geq u(\bar{a}) - u(0)$ , then under full clarity the whole budget is allocated and the allocation is efficient. This is strictly better for the principal than any outcome under divided agency unless the efficient allocation happens to be  $a_1 = a_2 = \frac{\bar{a}}{2}$ . And, even in this event, it is weakly better than any outcome under divided agency.

Now suppose  $\bar{B} < u(\bar{a}) - u(0)$ . Under full clarity the equilibrium allocations are  $(0, \hat{a}_2(0))$  and under divided agency the equilibrium allocations are  $(a_1^{NU}, a_2^{NU})$ .

I start by showing that  $\hat{a}_2(0) > a_2^{NU}$ .

**Lemma A.1** *If  $\bar{B} + u(0) < u(\bar{a})$ , then  $\hat{a}_2(0) > a_2^{NU}$ .*

**Proof.** First, suppose  $a_2^{NU} < \frac{\bar{a}}{2}$ . If  $a_2^{NU} \geq \hat{a}_2(0)$ , then we have

$$\begin{aligned} \bar{B} &\leq u(\bar{a}) - u(\bar{a} - a_2^{NU}) \\ &\leq u\left(\frac{\bar{a}}{2}\right) - u\left(\frac{\bar{a}}{2} - a_2^{NU}\right) \\ &= \underline{B}, \end{aligned}$$

where the first inequality is from the definition of  $\hat{a}_2(0)$  and the hypothesis that  $a_2^{NU} \geq \hat{a}_2(0)$ , the second inequality is from the weak concavity of  $u$ , and the equality is from the definition of  $a_2^{NU}$ . But this implies  $\bar{B} \leq \underline{B}$ , a contradiction.

Now suppose  $a_2^{NU} = \frac{\bar{a}}{2}$ . If  $a_2^{NU} \geq \hat{a}_2(0)$ , then we have

$$\begin{aligned} \bar{B} &\leq u(\bar{a}) - u\left(\frac{\bar{a}}{2}\right) \\ &\leq u\left(\frac{\bar{a}}{2}\right) - u(0) \\ &= \underline{B}, \end{aligned}$$

a contradiction. ■

No unified agency is preferred to full clarity if:

$$\gamma f(a_1^{NU}) + (1 - \gamma)(g(a_2^{NU}) - g(\hat{a}_2(0))) \geq 0.$$

Given that none of  $a_1^{NU}$ ,  $a_2^{NU}$ , or  $\hat{a}_2(0)$  is a function of  $\gamma$ , Lemma A.1 makes clear that this inequality holds at  $\gamma = 1$  and doesn't hold at  $\gamma = 0$ . Thus, it suffices to prove that the left-hand

side is monotone increasing in  $\gamma$ . Differentiating the left-hand side with respect to  $\gamma$  yields:

$$f(a_1^{NU}) + g(\hat{a}_2(0)) - g(a_2^{NU}),$$

which, by Lemma A.1, is obviously positive. ■

### Proof of Proposition 6.5

It suffices to show that no clarity is dominated by observable actions with divided agency. From Proposition 6.3 if  $\underline{B} + u(0) \geq u\left(\frac{\bar{a}}{2}\right)$ , the principal is clearly better off with observable actions, since she extracts  $\left(\frac{\bar{a}}{2}, \frac{\bar{a}}{2}\right)$ . Turn to the case where  $\underline{B} + u(0) < u\left(\frac{\bar{a}}{2}\right)$ . We must show that  $a_i^{NU} > a_i^{NC}$ . To see this, note that from the interiority of  $(a_1^{NC}, a_2^{NC})$  we know that  $p(a_1^{NC}, a_2^{NC})\underline{B} + u\left(\frac{\bar{a}}{2} - a_1^{NC}\right) \geq p(0, a_2^{NC})\underline{B} + u\left(\frac{\bar{a}}{2}\right)$ . Moreover, from Equation 7, we have that  $\underline{B} + u\left(\frac{\bar{a}}{2} - a_1^{NU}\right) = u\left(\frac{\bar{a}}{2}\right)$ . Given this, we have the following:

$$\begin{aligned} u\left(\frac{\bar{a}}{2}\right) - u\left(\frac{\bar{a}}{2} - a_1^{NC}\right) &\leq (p(a_1^{NC}, a_2^{NC}) - p(0, a_2^{NC}))\underline{B} \\ &< \underline{B} \\ &= u\left(\frac{\bar{a}}{2}\right) - u\left(\frac{\bar{a}}{2} - a_1^{NU}\right), \end{aligned}$$

which establishes the result. The same argument holds for  $a_2^{NU} > a_2^{NC}$ . ■

## References

- Alt, James E. and David Dreyer Lassen. 2006. “Fiscal Transparency, Political Parties and Debt in OECD Countries.” *European Economic Review*.
- Ashworth, Scott and Kenneth W. Shotts. 2010. “Does Informative Media Commentary Reduce Politicians’ Pander?” *Journal of Public Economics* 94(11–12):838–847.
- Austen-Smith, David and Jeffrey Banks. 1989. Electoral Accountability and Incumbency. In *Models of Strategic Choice in Politics*, ed. Peter C. Ordeshook. Ann Arbor: University of Michigan Press.
- Baron, David and David Besanko. 1992. “Information, Control, and Organizational Structure.” *Journal of Economics & Management Strategy* 1:237–275.

- Barro, Robert. 1973. "The Control of Politicians: An Economic Model." *Public Choice* 14:19–42.
- Berry, Christopher R. and Jacob E. Gersen. 2008. "The Unbundled Executive." *University of Chicago Law Review* 75:1385.
- Besley, Timothy and Stephen Coate. 2003. "Elected versus appointed regulators: Theory and evidence." *Journal of the European Economic Association* 1(5):1176–1206.
- Canes-Wrone, Brandice, Michael C. Herron and Kenneth W. Shotts. 2001. "Leadership and Pandering: A Theory of Executive Policymaking." *American Journal of Political Science* 45:532–550.
- Duch, Raymond M. and Randy Stevenson. 2007. "Context and Strategic Economic Voting: An Alternative to Clarity of Responsibility?" Working Paper.
- Ferejohn, John. 1986. "Incumbent Performance and Electoral Control." *Public Choice* 50:5–26.
- Fox, Justin. 2007. "Government Transparency and Policymaking." *Public Choice* 131(April):23–44.
- Fox, Justin and Kenneth W. Shotts. 2009. "Delegates or Trustees? A Theory of Political Accountability." *Journal of Politics* 71:1225–1237.
- Fox, Justin and Matthew C. Stephenson. 2011. "Judicial Review as a Response to Political Posturing." *American Political Science Review* 105(2):397–414.
- Fox, Justin and Richard van Weelden. 2012. "Costly Transparency." *Journal of Public Economics* 96(1-2):142150.
- Gailmard, Sean and John W. Patty. 2009. "Separation of Powers, Information, and Bureaucratic Structure." Berkeley Typescript.
- Gersen, Jacob E. 2010. "Unbundled Powers." *Virginia Law Review* 96:301–1965.
- Gilbert, Richard J. and Michael H. Riordan. 1995. "Regulating Complementary Products: A Comparative Institutional Analysis." *RAND Journal of Economics* 26:243–256.
- Hatfield, John William and Gerard Padró i Miquel. 2006. "Multitasking, Limited Liability, and Political Agency." Stanford Typescript.

- Holmström, Bengt. 1982. "Moral Hazard in Teams." *Bell Journal of Economics* 13(2):324–340.
- Holmström, Bengt and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7:24–52.
- Khalil, Fahad, Doyoung Kim and Dongsoo Shin. 2006. "Optimal Task Design: To Integrate or Separate Planning and Implementation." *Journal of Economics & Management Strategy* 15(Summer):457–478.
- Laffont, Jean-Jacques and Jean Tirole. 1988. "Repeated Auctions of Incentive contracts, Investment, and Bidding Parity with an Application to Takeovers." *RAND Journal of Economics* 19:516–537.
- Lewis-Beck, Michael. 1988. *Economics and Elections: The Major Western Democracies*. Ann Arbor: University of Michigan Press.
- Lewis, Tracey and David Sappington. 1997. "Ignorance in Agency Problems." *Journal of Economic Theory* 61:169–183.
- Maskin, Eric and Jean Tirole. 2004. "The Politician and the Judge: Accountability in Government." *American Economic Review* 94(4):1034–1054.
- McCubbins, Mathew and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science* 28(1):16–79.
- Nadeau, Richard, Richard G. Niemi and Antoine Yoshinaka. 2002. "A Cross-National Analysis of Economic Voting: Taking Account of the Political Context Across Time and Nations." *Electoral Studies* 21(3):403–42.
- Persson, Torsten, Gerard Roland and Guido Tabellini. 1997. "Separation of Powers and Political Accountability." *Quarterly Journal of Economics* 112:1163–1202.
- Persson, Torsten and Guido Tabellini. 2000. *Political Economics: Explaining Economic Policy*. Cambridge: MIT Press.
- Powell, G. Bingham. 2000. *Elections as Instruments of Democracy: Majoritarian and Proportional Visions*.

- Powell, G. Bingham and Guy D. Whitten. 1993. "A Cross-National Analysis of Economic Voting: Taking Account of the Political Context." *American Journal of Political Science* 37(May):391–414.
- Riordan, Michael and David Sappington. 1987. "Information, Incentives, and Organizational Mode." *Quarterly Journal of Economics* 102:243–263.
- Royed, Terry J., Kevin M. Leyden and Stephen A. Borrelli. 2000. "Is 'Clarity of Responsibility' Important for Economic Voting? Revisiting Powell and Whitten's Hypothesis." *British Journal of Political Science* 30(4):669–698.
- Samuels, David. 2004. "Presidentialism and Accountability for the Economy in Comparative Perspective." *American Political Science Review* 98(3):425–436.
- Seabright, Paul. 1996. "Accountability and Decentralisation in Government: An Incomplete Contracts Model." *European Economic Review* 40:61–89.
- Shi, Min and Jakob Svensson. 2006. "Political Budget Cycles: Do They Differ Across Countries and Why?" *Journal of Public Economics* 90(8-9):1367–1389.
- Shotts, Kenneth W. and Alan Wiseman. 2010. "The Politics of Investigations and Regulatory Enforcement by Independent Agents and Cabinet Appointees." *Journal of Politics* 72(1):209–226.
- Stephenson, Matthew C. 2008. "Optimal Political Control of the Bureaucracy." *Michigan Law Review* 107(53):54–110.
- Tavits, Margit. 2007. "Clarity of Responsibility and Corruption." *American Journal of Political Science* 51(1).
- Ting, Michael M. 2003. "A Strategic Theory of Bureaucratic Redundancy." *American Journal of Political Science* 47(2):274–292.
- Weingast, Barry R., Kenneth A. Shepsle and Christopher Johnsen. 1981. "The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics." *Journal of Political Economy* 89(4):642–664.

Whitten, Guy D. and Harvey D. Palmer. 1999. "Cross-National Analysis of Economic Voting."  
*Electoral Studies* 18(1):49–67.

## Appendix B Supplemental Material 1: Simultaneous Policy Making

Consider the game with one policy-making stage, one agent, and observable actions. A strategy for the principal is a pair,  $(m_1, \rho)$ , where  $m_1(\cdot, \cdot, \cdot) \in \mathcal{R}$  is an initial declaration of a retention rule and  $\rho(m_1, a_1, a_2, O)(\cdot, \cdot, \cdot) \in \mathcal{R}$  is the retention rule that will be used following a history  $(m_1, a_1, a_2, O)$ . A strategy for the agent is a pair  $(a_1, a_2)$  that maps initial declarations,  $m_1$ , into action choices. We denote equilibrium choices with the superscript  $C_{si}$  (for *clarity* with *simultaneous choices*).

Lemma C.5 in Appendix C gives necessary and sufficient conditions for a strategy profile in this game to be an equilibrium. In such an equilibrium, the principal declares a rule that, when best responded to, maximizes her expected utility, and she sticks with that rule. Given this, what happens in equilibrium?

The agent's payoff from an allocation  $(a_1, a_2)$  and being retained with certainty is  $\bar{B} + u(\bar{a} - a_1 - a_2)$ . Her payoff from allocating nothing and not being retained is  $u(\bar{a})$ . Define  $\hat{a}$  as the largest level of total spending that satisfies:

$$\bar{B} + u(\bar{a} - \hat{a}) \leq u(\bar{a}).$$

The principal can induce the agent to allocate any amount  $a_1 + a_2 \leq \hat{a}$  by retaining if and only if the agent allocates precisely that amount. Moreover, given any total level of spending, the agent is indifferent as to its division between  $a_1$  and  $a_2$ . So the principal can induce an efficient division of  $\hat{a}$  by choosing a rule that retains the agent if and only if the agent chooses the allocation

$$(a_1^{C_{si}}, a_2^{C_{si}}) = (a_1^{\text{efficient}(\hat{a})}, a_2^{\text{efficient}(\hat{a})}). \quad (8)$$

This requires that  $p_1(a_1^{C_{si}}, a_2^{C_{si}}) = p_2(a_1^{C_{si}}, a_2^{C_{si}})$ . Since  $p$  is increasing and concave in both of its arguments, there is a unique such allocation.

Finally, note that it is impossible for the principal to do any better by allowing her retention rule to condition on the outcome. Doing so simply introduces a stochastic element into the agent's payoffs, decreasing the power of his incentives. Given this, we have the following result.

**Proposition B.1** *Suppose there is full clarity of responsibility and  $a_1$  and  $a_2$  are chosen simultaneously. Any equilibrium has the property that, on the equilibrium path the agent chooses  $(a_1^{C_{si}}, a_2^{C_{si}})$  as defined by equation 8. Moreover, if  $\bar{B} + u(0) \geq u(\bar{a})$ , the agent is allocating the full budget.*

**Proof.** Follows from Lemma C.5 and the argument in the text. ■

From the principal's perspective the equilibrium allocation under full clarity with static policy making cannot be improved upon by eliminating any aspect of clarity of responsibility. In particular, the principal induces the agent to either allocate the full budget or to allocate a level of resources that just satisfies the agent's participation constraint. Moreover, the principal induces the agent to divide these resources efficiently. This case, thus, comports with the intuition that clarity of responsibility facilitates accountability.

## Appendix C Supplemental Material 2: Necessary and Sufficient Conditions for SPNE that are Robust to Small Revision Costs

**Lemma C.1** *Consider a game with two agents and no action observability. An associated SPNE,  $s^*$ , is robust to small revision costs if and only if*

$$(i) \rho^*(m_1, m_2, O)(\cdot) = m_2(\cdot) \text{ for all } (m_1, m_2, O)$$

$$(ii) a_2^*(m_1, m_2) \in \arg \max_{a_2 \in [0, \frac{\bar{a}}{2}]} \mathbb{E}[m_2^2(O)\underline{B} + u(\frac{\bar{a}}{2} - a_2) | a_1^*(m_1)] \text{ for all } (m_1, m_2);$$

$$(iii) m_2^{1,*}(m_1)(\cdot) = m_1^1(\cdot) \text{ for all } (m_1);$$

$$(iv) m_2^{2,*}(m_1)(\cdot) \in \arg \max_{m_2^2 \in \mathcal{R}} p(a_1^*(m_1), a_2^*(m_1, (m_1^1, m_2^2))) \text{ for all } m_1;$$

$$(v) a_1^*(m_1) \in \arg \max_{a_1 \in [0, \frac{\bar{a}}{2}]} \mathbb{E}[m_1^1(O)\underline{B} + u(\frac{\bar{a}}{2} - a_1) | a_2^*(m_2^*(m_1))] \text{ for all } m_1;$$

$$(vi) m_1^*(\cdot) \in \arg \max_{m_1(\cdot) \in \mathcal{R}} p(a_1^*(m_1), a_2^*(m_1, m_2^*(m_1))).$$

**Proof. Necessity:** Fix an  $s^*$  that is robust to small revision costs.

Consider the retention stage. Fix an  $\epsilon$ . At this stage, the agent's actions are already taken. Hence, the principal's action at this stage has no effect on the choice of agent actions. However, if  $\rho^\epsilon(m_1, m_2, O) \neq m_2$ , the principal suffers a cost  $\epsilon$ . Hence, sequential rationality requires  $\rho^\epsilon(m_1, m_2, O) = m_2$  for all  $(m_1, m_2, O)$ . This implies that  $\rho^\epsilon(m_1, m_2, O) = \rho^*(m_1, m_2, O)$  for all  $(m_1, m_2, O)$ , which establishes point 1.

Next consider agent 2's strategy in the second policy-making stage. For any  $\epsilon$ , the agent infers that the principal will choose  $\rho^\epsilon(m_1, m_2, O)(\cdot) = m_2(\cdot)$ . Hence, the agent chooses  $a_2^\epsilon$  to maximize his expected utility given the retention rule  $m_2(\cdot)$  and his beliefs about  $a_1$ , which must be correct in equilibrium. This implies that  $a_2^\epsilon(m_1, m_2) = a_2^*(m_1, m_2)$  for all  $(m_1, m_2)$ , which establishes point 2.

Next consider the principal's strategy in the second policy-making stage. First consider  $m_2^1$ . Fix an  $\epsilon$ . At this stage, the first agent's actions are already taken. Hence, the principal's choice of  $m_2^1$  has no effect on the choice of agent actions. However, if  $m_2^1(m_1, a_1) \neq m_1^1$ , the principal suffers a cost  $\epsilon$ . Hence, sequential rationality requires  $m_2^1(m_1, a_1) = m_1^1$  for all  $(m_1, a_1)$ . This establishes point 3.

Now consider  $m_2^2$ . Suppose the principal believes agent 1 took action  $\tilde{a}_1$ . For any  $\epsilon$ , define  $\hat{m}_2^2(\tilde{a}_1) = \arg \max_{m_2^2 \in \mathcal{R}} p(\tilde{a}_1, a_2^\epsilon(m_1, (m_1^1, m_2^2)))$ . For any  $\epsilon$ , given that the action  $a_1$  is already taken, the principal chooses  $m_2^2$  to maximize her expected utility net of any revision costs. Thus, she either chooses  $m_2^2 = m_1^2$  or she chooses  $m_2^2 = \hat{m}_2^2(\tilde{a}_1)$ . By definition, if  $\hat{m}_2^2(\tilde{a}_1) \neq m_1^2$ , then choosing  $m_2^2 = \hat{m}_2^2(\tilde{a}_1)$  yields a higher choice of  $a_2$ . Hence, for each  $(m_1, \tilde{a}_1)$ ,  $m_2^2$  equals  $\hat{m}_2^2(\tilde{a}_1)$  if  $\epsilon$  is sufficiently small. Call the minimal  $\epsilon$ ,  $\bar{\epsilon}(m_1, \tilde{a}_1)$ . It suffices to show that, given that it converges (which we know by hypothesis), the sequence  $\{m_2^{2,\epsilon}\}$  converges to  $\hat{m}_2^2(\tilde{a}_1)$ . This is obvious, since we have shown that  $m_2^{2,\epsilon}(m_1, \tilde{a}_1) = \hat{m}_2^2(\tilde{a}_1)$  for all  $\epsilon < \bar{\epsilon}(m_1, \tilde{a}_1)$  for all  $(m_1, \tilde{a}_1)$ . Moreover, in equilibrium beliefs must be correct, so  $\tilde{a}_1 = a_1^*(m_1)$ . This establishes point 4.

Next consider agent 1's strategy in the first policy-making stage. Fix an  $\epsilon$ . Define  $m_2^\epsilon(m_1)$  as whichever of  $m_1^2$  and  $\hat{m}_2^2(\tilde{a}_1)$  maximize the principal's expected utility at the second policy-making stage. By backward induction, the agent knows that for any  $m_1$ , the principal will choose  $m_2^\epsilon(m_1)$  and the second agent will then best respond with  $a_2^*(m_1, m_2^\epsilon(m_1))$ , since  $a_2^\epsilon = a_2^*$  for all  $\epsilon$ . Hence, the agent chooses  $a_1^\epsilon(m_1) \in \arg \max_{a_1 \in [0, \frac{\bar{a}}{2}]} \mathbb{E}[m_2^\epsilon(m_1)(O)\underline{B} + u(\frac{\bar{a}}{2} - a_1) | a_2^*(m_1, m_2^\epsilon(m_1))]$ . By the argument above,  $m_2^\epsilon$  converges to  $m_2^*$ , so  $a_1^\epsilon$  converges to  $a_1^*$ , which establishes point 5.

Finally, consider the principals' strategy in the first policy-making stage. Clearly, for any  $\epsilon$ , in any equilibrium,  $m_1^\epsilon$  must satisfy the condition in point 6, substituting  $\epsilon$  for  $*$ . Further, the preceding arguments show that  $a_2^\epsilon$  converges to  $a_2^*$ ,  $a_1^\epsilon$  converges to  $a_1^*$ , and  $m_2^\epsilon$  converges to  $m_2^*$ . Hence, it is obvious that  $m_1^\epsilon$  converges to  $m_1^*$ , which establishes point 6.

**Sufficiency:** Fix an  $s^*$  satisfying 1–6.

From the necessity proof, with the exception of  $m_1$  and  $m_2^{2,*}$ , the strategies in  $s^*$  are identical to the strategies in a SPNE of a game with any  $\epsilon$ . Thus, all we need to do is show that point 6 implies convergence of  $m_1^\epsilon$  to  $m_1^*$  and that point 4 implies convergence of  $m_2^{2,\epsilon}$  to  $m_2^{2,*}$ .

First consider  $m_2^2$ . Fix a  $\delta > 0$ . Now we must show that there exists an  $\epsilon$  such that  $|m_2^{2,\epsilon}(m_1) - m_2^{2,*}(m_1)| < \delta$ . By point 3,  $m_2^{2,*}(m_1)(\cdot) = \hat{m}_2(a_1^*(m_1))(\cdot)$  for all  $m_1$ . Thus, it suffices to show there exists an  $\epsilon$  such that  $|m_2^{2,\epsilon}(m_1) - \hat{m}_2^2(a_1^*(m_1))| < \delta$ .

Suppose for some  $\epsilon$  and some  $m_1$ ,  $m_2^{2,\epsilon}(m_1) = m_1^2 \neq \hat{m}_2^2(a_1^*(m_1))$ . Then this implies that the principal's expected utility is  $U_P(a_1^*(m_1), a_2^*(m_1, m_2^{2,\epsilon}(m_1)))$  which is less than  $U_P(a_1^*(m_1), a_2^*(\hat{m}_2(a_1^*(m_1))))$ . Call the difference  $\gamma$ . Now select a new  $\epsilon' < \gamma$ , and  $m_2^{2,\epsilon'}(m_1)(\cdot) = \hat{m}_2^2(a_1^*(m_1))(\cdot)$ .

The fact that  $m_1$  converges now follows immediately from the fact that  $a_1^\epsilon$  converges to  $a_1^*$ ,  $a_2^\epsilon$  converges to  $a_2^*$  and  $m_2^\epsilon$  converges to  $m_2^*$ . ■

**Lemma C.2** *Consider a game with one agent and observable actions. An associated SPNE,  $s^*$ , is robust to small revision costs if and only if*

- (i)  $\rho^*(m_1, m_2, a_1, a_2)(\cdot, \cdot) = m_2(\cdot, \cdot)$  for all  $(m_1, m_2, a_1, a_2)$ ;
- (ii)  $a_2^*(m_1, m_2, a_1) \in \arg \max_{a_2 \in [0, \bar{a} - a_1]} m_2(a_1, a_2) \bar{B} + u(\bar{a} - a_1 - a_2)$ , for all  $(m_1, m_2, a_1)$ ;
- (iii)  $m_2^*(m_1, a_1)(\cdot, \cdot) \in \arg \max_{m_2 \in \mathcal{R}} p(a_1, a_2^*(m_1, m_2, a_1))$ , for all  $(m_1, a_1)$ ;
- (iv)  $a_1^*(m_1) \in \arg \max_{a_1 \in [0, \bar{a}]} m_2^*(m_1, a_1)(a_1, a_2^*(m_1, m_2^*(m_1, a_1), a_1)) \bar{B} + u(\bar{a} - a_1 - a_2^*(m_1, m_2^*(m_1, a_1), a_1))$ ,  
for all  $m_1$ ;
- (v)  $m_1^* \in \arg \max_{m_1 \in \mathcal{R}} p(a_1^*(m_1), a_2^*(m_1, m_2^*(a_1^*(m_1), m_1), a_1^*(m_1)))$ .

**Proof. Necessity:** Fix an  $s^*$  that is robust to small revision costs.

Consider the retention stage. Fix an  $\epsilon$ . At this stage, the agent's actions are already taken. The principal's action at this stage has no effect on the choice of agent actions. However, if  $\rho^\epsilon(m_1, m_2, a_1, a_2) \neq m_2$ , the principal suffers a cost  $\epsilon$ . Hence, sequential rationality requires  $\rho^\epsilon(m_1, m_2, a_1, a_2) = m_2$  for all  $(m_1, m_2, a_1, a_2)$ . Since this is true for all  $\epsilon$  and the sequence converges, this establishes point 1.

Next consider the agent's strategy in the second policy-making stage. For any  $\epsilon$ , the agent infers that the principal will choose  $\rho^\epsilon(m_1, m_2, a_1, a_2) = m_2$ . Hence, the agent chooses  $a_2^\epsilon$  to

maximize his expected utility given the retention rule  $m_2$  and the history. That is, for any  $\epsilon$ ,  $a_2^\epsilon(m_1, m_2, a_1) = a_2^*(m_1, m_2, a_1)$  for any  $(m_1, m_2, a_1)$ . This establishes point 2.

Next consider the principal's strategy in the second policy-making stage. Define  $\hat{m}_2(a_1) = \arg \max_{m_2 \in \mathcal{R}} p(a_1, a_2^*(m_1, m_2, a_1))$ . For any  $\epsilon > 0$ , given that the action  $a_1$  is already taken, the principal chooses  $m_2$  to maximize her utility net of any revision costs. Thus, she either chooses  $m_2 = m_1$  or she chooses  $m_2 = \hat{m}_2(a_1)$ . For each history  $(m_1, a_1)$  with  $m_1 \neq \hat{m}_2(a_1)$ , she chooses  $m_2 = \hat{m}_2(a_1)$  for  $\epsilon$  sufficiently small, in particular, she does so if  $\max_{m_2 \in \mathcal{R}} p(a_1, a_2^\epsilon(m_1, m_2, a_1)) - p(a_1, a_2^\epsilon(m_1, m_1, a_1)) \geq \epsilon$ . Label with  $\bar{\epsilon}(m_1, a_1)$ , the  $\epsilon$  that makes this hold with equality. Recall, from the previous paragraph, that  $a_2^\epsilon(m_1, m_2, a_1) = a_2^*(m_1, m_2, a_1)$  for all  $(m_1, m_2, a_1)$ . Hence,  $\hat{m}_2(a_1) = m_2^*(m_1, m_2, a_1)$  for all  $(m_1, m_2, a_1)$ . Thus, it suffices to show that, given that it converges (which we know by hypothesis), the sequence  $\{m_2^\epsilon\}$  converges to  $\hat{m}_2(a_1)$ . This is clearly true, since  $m_2^\epsilon(m_1, a_1) = \hat{m}_2(a_1)$  for all  $\epsilon < \bar{\epsilon}(m_1, a_1)$  for all  $(m_1, a_1)$ . This establishes point 3.

Next consider the agent's strategy in the first policy-making stage. Fix an  $\epsilon > 0$ . Define  $m_2^\epsilon(m_1, a_1)$  as whichever of  $m_1$  and  $\hat{m}_2(a_1)$  maximizes the principal's expected utility at the second policy-making stage. By backward induction, the agent knows that for any  $(m_1, a_1)$ , the principal will choose  $m_2^\epsilon(m_1, a_1)$  and the agent himself will then best respond with  $a_2^*(m_1, m_2^\epsilon(m_1, a_1), a_1)$ . Hence, the agent chooses

$$a_1^\epsilon(m_1) \in \arg \max_{a_1 \in [0, \bar{a}]} \mathbb{E}[m_2^\epsilon(m_1, a_1)(a_1, a_2^*(a_1, m_1, m_2^\epsilon(m_1, a_1)))\bar{B} + u(\bar{a} - a_1 - a_2^*(m_1, m_2^\epsilon(m_1, a_1)), a_1)].$$

We need to show that, given that it converges, the sequence  $\{a_1^\epsilon(m_1)\}$  converges to  $a_1^*(m_1)$ . Given an  $\epsilon$ , for any  $(m_1, a_1)$ , the principal chooses  $m_2^\epsilon(m_1, a_1)$  either equal to  $m_1$  or to  $m_2^*(m_1, a_1)$ . Suppose the agent anticipates that for his choice of  $a_1^\epsilon(m_1)$ , we have  $m_2^\epsilon(m_1, a_1^\epsilon(m_1)) = m_2^*(m_1, a_1^\epsilon(m_1))$ . Then  $a_1^\epsilon(m_1)$  must be a best response to  $m_2^*(m_1, a_1^\epsilon(m_1))$ , which implies  $a_1^\epsilon(m_1) = a_1^*(m_1)$ . Now suppose the agent anticipates that for his choice of  $a_1^\epsilon(m_1)$ , we have  $m_2^\epsilon(m_1, a_1^\epsilon(m_1)) = m_1 \neq m_2^*(m_1, a_1^\epsilon(m_1))$ . By the fact that  $m_2^\epsilon \neq m_2^*$ , there exists a second period retention rule that would yield a higher second period action. So for  $\epsilon$  low enough (in particular,  $\epsilon' < \bar{\epsilon}(m_1, a_1^\epsilon(m_1))$ ) we have that  $m_2^{\epsilon'}(m_1, a_1^\epsilon(m_1)) = m_2^*(m_1, a_1^\epsilon(m_1))$ . Now, since  $a_1^\epsilon(m_1) = a_1^{\epsilon'}(m_1) = a_1^*(m_1)$ , this takes us back to the first case and establishes point 4.

Finally, consider the principal's strategy in the first policy-making stage. For any  $(m_1, a_1)$ ,  $m_2^\epsilon(m_1, a_1) = m_2^*(m_1, a_1)$  for  $\epsilon \leq \bar{\epsilon}(m_1, a_1)$ . Moreover, from point 3,  $m_2^*(m_1, a_1)$  is constant in  $m_1$ . Hence, for  $\epsilon$  sufficiently small, backward induction implies that  $a_1^\epsilon(m_1)$  is unaffected

by the choice of  $m_1$ . Consider such an  $\epsilon$ . Suppose  $m_1^\epsilon \neq m_2^\epsilon(m_1^\epsilon, a_1^\epsilon(m_1^\epsilon))$ . Consider a deviation to a new  $m_1 = m_2^\epsilon(m_1^\epsilon, a_1^\epsilon(m_1^\epsilon))$ . This makes the principal strictly better off. To see this, note that as already argued,  $a_1^\epsilon$  is left unchanged. But the principal avoids revising the rule and bearing cost  $\epsilon$ . Thus, for  $\epsilon$  sufficiently small, sequential rationality implies that  $m_1^\epsilon \in \arg \max_{r \in \mathcal{R}} p(a_1^\epsilon(r), a_2^\epsilon(r, m_2^*(a_1^\epsilon(r), r), a_1^\epsilon(r)))$ . Now note that  $a_1^\epsilon = a_1^*$  and  $a_2^\epsilon = a_2^*$ . This establishes point 5.

**Sufficiency:** Fix an  $s^*$  satisfying 1–5.

From the necessity proof, the agent's strategy in  $s^*$  is identical to his strategy in a SPNE of a game with any  $\epsilon$ . Thus, all we need to show is convergence of the principal's strategy.

First, consider whether  $m_2^\epsilon$  converges to  $m_2^*$ . Fix a  $\delta > 0$ . We must show that there exists an  $\epsilon$  such that  $|m_2^\epsilon(m_1, a_1) - m_2^*(m_1, a_1)| < \delta$ , for all  $(m_1, a_1)$ . By point 3,  $m_2^*(m_1, a_1) = \hat{m}_2(a_1)$  for all  $(m_1, a_1)$ . Thus, it suffices to show there exists an  $\epsilon$  such that  $|m_2^\epsilon(m_1, a_1) - \hat{m}_2(a_1)| < \delta$ . This is true for all  $\epsilon < \bar{\epsilon}(m_1, a_1)$ , since for such  $\epsilon$ ,  $m_2^\epsilon(m_1, a_1) = \hat{m}_2(a_1)$ .

Finally, consider whether  $m_1^\epsilon$  converges to  $m_1^*$ . Fix a  $\delta > 0$ . By sequential rationality, for any  $\epsilon$ ,  $m_1^\epsilon \in \arg \max_{r \in \mathcal{R}} p(a_1^*(r), a_2^*(r, m_2^\epsilon(a_1^*(r), r), a_1^*(r)))$ . And, as we saw above, for  $\epsilon$  sufficiently small,  $m_2^\epsilon = m_2^*$ . Hence, for those same  $\epsilon$ ,  $m_1^\epsilon = m_1^*$ . ■

**Lemma C.3** *Consider a game with one agent and no action observability. An associated SPNE,  $s^*$ , is robust to small revision costs if and only if*

- (i)  $\rho^*(m_1, m_2, O)(\cdot) = m_2(\cdot)$  for all  $(m_1, m_2, O)$
- (ii)  $a_2^*(m_1, m_2, a_1) \in \arg \max_{a_2 \in [0, \bar{a} - a_1]} \mathbb{E}[m_2(O)\bar{B} + u(\bar{a} - a_1 - a_2)]$  for all  $(m_1, m_2, a_1)$ ;
- (iii)  $m_2^*(m_1)(\cdot) \in \arg \max_{m_2 \in \mathcal{R}} p(a_1^*(m_1), a_2^*(m_1, m_2, a_1^*(m_1)))$  for all  $m_1$ ;
- (iv)  $a_1^*(m_1) \in \arg \max_{a_1 \in [0, \bar{a}]} \mathbb{E}[m_2^*(m_1)(O)\bar{B} + u(\bar{a} - a_1 - a_2^*(m_1, m_2^*(m_1), a_1))]$  for all  $m_1$ ;
- (v)  $m_1^*(\cdot) = m_2^*(m_1^*)(\cdot)$ .

**Proof. Necessity:** Fix an  $s^*$  that is robust to small revision costs.

Consider the retention stage. Fix an  $\epsilon$ . At this stage, the agent's actions are already taken. Hence, the principal's action at this stage has no effect on the choice of agent actions. However, if  $\rho^\epsilon(m_1, m_2, O) \neq m_2$ , the principal suffers a cost  $\epsilon$ . Hence, sequential rationality requires

$\rho^\epsilon(m_1, m_2, O) = m_2$  for all  $(m_1, m_2, O)$ . Since this is true for all  $\epsilon$  and the sequence converges, this establishes point 1.

Next consider the agent's strategy in the second policy-making stage. For any  $\epsilon$ , the agent infers that the principal will choose  $\rho^\epsilon(m_1, m_2, O)(\cdot) = m_2(\cdot)$ . Hence, the agent chooses  $a_2^\epsilon$  to maximize his expected utility given the retention rule  $m_2(\cdot)$  and the history. Since this is true for all  $\epsilon$  and the sequence converges, this establishes point 2.

Next consider the principal's strategy in the second policy-making stage. Suppose the principal believes the agent took action  $\tilde{a}_1$ . For any  $\epsilon$ , define  $\hat{m}_2(\tilde{a}_1) = \arg \max_{m_2 \in \mathcal{R}} p(\tilde{a}_1, a_2^\epsilon(\tilde{a}_1, m_1, m_2))$ . For any  $\epsilon$ , given that the action  $a_1$  is already taken, the principal chooses  $m_2$  to maximize her expected utility net of any revision costs. Thus, she either chooses  $m_2 = m_1$  or she chooses  $m_2 = \hat{m}_2(\tilde{a}_1)$ . Notice for each belief  $\tilde{a}_1$ , she only chooses  $\hat{m}_2(\tilde{a}_1)$  for  $\epsilon$  sufficiently small. Call the minimal  $\epsilon$ ,  $\bar{\epsilon}(m_1, \tilde{a}_1)$ . It suffices to show that, given that it converges (which we know by hypothesis), the sequence  $\{m_2^\epsilon\}$  converges to  $\hat{m}_2(\tilde{a}_1)$ . This is obvious, since we have shown that  $m_2^\epsilon(m_1, \tilde{a}_1) = \hat{m}_2(\tilde{a}_1)$  for all  $\epsilon < \bar{\epsilon}(m_1, \tilde{a}_1)$ . In equilibrium beliefs must be correct, so  $\tilde{a}_1 = a_1^*(m_1)$ . This establishes point 3.

Next consider the agent's strategy in the first policy-making stage. Fix an  $\epsilon > 0$ . Define  $m_2^\epsilon(m_1)$  as whichever of  $m_1$  and  $\hat{m}_2(\tilde{a}_1)$  maximize the principals expected utility at the second policy-making stage. By backward induction, the agent knows that for any  $m_1$ , the principal will choose  $m_2^*(m_1)$  and the agent himself will then best respond with  $a_2^*(m_1, m_2^*(m_1), a_1)$ . Hence, the agent chooses

$$a_1^\epsilon(m_1) \in \arg \max_{a_1 \in [0, \bar{a}]} \mathbb{E}[m_2^*(m_1)(a_1, a_2^*(a_1, m_1, m_2^*(m_1)), O) \bar{B} + u(\bar{a} - a_1 - a_2^*(a_1, m_1, m_2^*(m_1)))].$$

Since this is true for all  $\epsilon$  and the sequence converges by hypothesis, this establishes point 4.

Finally, consider the principals' strategy in the first policy-making stage. Fix an  $\epsilon$ . For any  $m_1$ , the agent chooses  $a_1^\epsilon(m_1)$  as a best response to  $m_2^\epsilon(m_1)$ . Suppose  $m_2^\epsilon(m_1^\epsilon) \neq m_1^\epsilon$ . Consider a deviation to a new  $m_1 = m_2^\epsilon(m_1^\epsilon)$ . This makes the principal strictly better off. To see this, note that  $a_1^\epsilon$  is left unchanged but the principal either avoids revising the rule and bearing cost  $\epsilon$  or not revising the rule and having a lower  $a_2$ . Hence, this is a profitable deviation, which establishes point 5.

**Sufficiency:** Fix an  $s^*$  satisfying 1–5.

From the necessity proof, with the exception of  $m_2^*$ , the strategies in  $s^*$  are identical to the

strategies in a SPNE of a game with any  $\epsilon$ . Thus, all we need to do is show that point 3 implies convergence of  $m_2^\epsilon$  to  $m_2^*$

Fix a  $\delta > 0$ . Now we must show that there exists an  $\epsilon$  such that  $|m_2^\epsilon(m_1) - m_2^*(m_1)| < \delta$ . By point 3,  $m_2^*(m_1)(\cdot) = \hat{m}_2(a_1^*(m_1))(\cdot)$  for all  $m_1$ . Thus, it suffices to show there exists an  $\epsilon$  such that  $|m_2^\epsilon(m_1, a_1) - \hat{m}_2(a_1)| < \delta$ .

Suppose for some  $\epsilon$  and some  $m_1$ ,  $m_2^\epsilon(m_1) = m_1 \neq \hat{m}_2(a_1^*(m_1))$ . Then this implies that the principal's expected utility is  $U_P(a_1^*(m_1), a_2^*(m_1, a_1^*(m_1)))$  which is less than  $U_P(a_1^*(m_1), a_2^*(\hat{m}_2(a_1^*(m_1)), a_1^*(m_1)))$ . Call the difference  $\gamma$ . Now select a new  $\epsilon' < \gamma$ , and  $m_2^{\epsilon'}(m_1)(\cdot) = \hat{m}_2(a_1^*(m_1))(\cdot)$ . This completes the proof. ■

**Lemma C.4** *Consider a game with two agents and action observability. An associated SPNE,  $s^*$ , is robust to small revision costs if and only if*

- (i)  $\rho^*(m_1, m_2, a_1, a_2, O) = m_2$  for all  $(m_1, m_2, a_1, a_2, O)$ ;
- (ii)  $a_2^*(m_1, m_2, a_1) \in \arg \max_{a_2 \in [0, \frac{\bar{a}}{2}]} \mathbb{E}[m_2^2(a_1, a_2, O)\underline{B} + u(\frac{\bar{a}}{2} - a_2)]$ , for all  $(m_1, m_2, a_1)$ ;
- (iii)  $m_2^{1,*}(m_1, a_1) = m_1^1$  for all  $(m_1, a_1)$ .
- (iv)  $m_2^{2,*}(a_1, m_1) \in \arg \max_{m_2^2 \in \mathcal{R}} p(a_1, a_2^*(a_1, m_1, (m_1^1, m_2^2)))$ , for all  $(m_1, a_1)$ ;
- (v)  $a_1^*(m_1) \in \arg \max_{a_1 \in [0, \frac{\bar{a}}{2}]} \mathbb{E}[m_1^1(a_1, a_2^*(a_1, m_1, (m_1^1, m_2^{2,*}(a_1, m_1))), O)\underline{B} + u(\frac{\bar{a}}{2} - a_1)]$ , for all  $m_1$ ;
- (vi)  $m_1^* \in \arg \max_{m_1 \in \mathcal{R}} p(a_1^*(m_1), a_2^*(m_1, (m_1^1, m_2^{2,*}(a_1^*(m_1), m_1))), a_1^*(m_1))$ .

**Proof. Necessity:** Fix an  $s^*$  that is robust to small revision costs.

Consider the retention stage. Fix an  $\epsilon$ . At this stage, the agent's actions are already taken. Hence, the principal's action at this stage has no effect on the choice of agent actions. However, if for any  $\rho^{\epsilon,i}(m_1, m_2, a_1, a_2, O) \neq m_2^i$ , the principal suffers a cost  $\epsilon$ . Hence, sequential rationality requires  $\rho^\epsilon(m_1, m_2, a_1, a_2, O) = m_2$  for all  $(m_1, m_2, a_1, a_2, O)$ , which implies that  $\rho^\epsilon(m_1, m_2, a_1, a_2, O) = \rho^*(m_1, m_2, a_1, a_2, O)$  for all  $(m_1, m_2, a_1, a_2, O)$ , establishing point 1.

Next consider agent 2's strategy in the second policy-making stage. For any  $\epsilon$ , the agent infers that the principal will choose  $\rho^\epsilon(m_1, m_2, a_1, a_2) = m_2$ . Hence, the agent chooses  $a_2^\epsilon$  to maximize

his expected utility given the retention rule  $m_2^2$  and the history, which implies that for any history  $a_2^\epsilon = a_2^*$ , which establishes point 2.

Next consider the principal's strategy in the second policy-making stage. First, consider  $m_2^1$ . Fix an  $\epsilon$ . At this stage, the first agent's actions are already taken. Hence, the principal's choice of  $m_2^1$  has no effect on the choice of agent actions. However, if  $m_2^1(m_1, a_1) \neq m_1^1$ , the principal suffers a cost  $\epsilon$ . Hence, sequential rationality requires  $m_2^1(m_1, a_1) = m_1^1$  for all  $(m_1, a_1)$ . Since this is true for all  $\epsilon$  and the sequence converges, this establishes point 3.

Now consider  $m_2^2$ . Define  $\hat{m}_2^2(a_1) = \arg \max_{m_2^2 \in \mathcal{R}} p(a_1, a_2^*(a_1, m_1, (m_1^1, m_2^2)))$ . For any  $\epsilon$ , given that the action  $a_1$  is already taken, the principal chooses  $m_2^2$  to maximize her utility net of any revision costs. Since  $a_2^\epsilon = a_2^*$ , this implies that she either chooses  $m_2^2 = m_1^1$  or she chooses  $m_2^2 = \hat{m}_2^2(a_1)$ . By definition, if  $\hat{m}_2^2(a_1) \neq m_1^1$ , then choosing  $m_2^2 = \hat{m}_2^2(a_1)$  yields a higher choice of  $a_2$ . Hence, for each  $(m_1, a_1)$   $m_2^2$  equals  $\hat{m}_2^2(a_1)$  if  $\epsilon$  is sufficiently small. Call the minimal  $\epsilon$ ,  $\bar{\epsilon}(m_1, a_1)$ . It suffices to show that, given that it converges (which we know by hypothesis), the sequence  $\{m_2^{2,\epsilon}\}$  converges to  $\hat{m}_2^2(a_1)$ . This is obvious, since we have shown that  $m_2^{2,\epsilon}(m_1, a_1) = \hat{m}_2^2(a_1)$  for all  $\epsilon < \bar{\epsilon}(m_1, a_1)$  for all  $(m_1, a_1)$ . This establishes point 4.

Next consider agent 1's strategy in first policy-making stage. Fix an  $\epsilon$ . By backward induction, the agent knows that for any  $(m_1, a_1)$ , the principal will choose  $\rho^{\epsilon,1}(m_1, m_2, a_1, a_2, O) = m_1^1$ . Hence, the agent chooses

$a_1^\epsilon(m_1) \in \arg \max_{a_1 \in [0, \frac{\bar{a}}{2}]} \mathbb{E}[m_1^1(a_1, a_2^*(a_1, m_1, (m_1^1, m_2^{2,\epsilon}(m_1, a_1))), O)B + u(\frac{\bar{a}}{2} - a_1)]$ . By the paragraph above, for all  $(m_1, a_1)$ ,  $m_2^{2,\epsilon}(m_1, a_1) = m_2^{2,*}(m_1, a_1)$  for  $\epsilon$  sufficiently small. Thus, given that it converges,  $a_1^\epsilon(m_1)$  must converge to  $a_1^*(m_1)$ , which establishes point 5.

Finally, consider the principals' strategy in the first policy-making stage. Clearly, for any  $\epsilon$ , in any equilibrium,  $m_1^\epsilon$  must satisfy the condition in point 6, substituting  $\epsilon$  for  $*$ . Further, the preceding arguments show that  $a_2^\epsilon = a_2^*$ , and that  $a_1^\epsilon$  converges to  $a_1^*$ , and  $m_2^\epsilon$  converges to  $m_2^*$ . Hence, it is obvious that, given that it converges,  $m_1^\epsilon$  converges to  $m_1^*$ , which establishes point 6.

**Sufficiency:** Fix an  $s^*$  satisfying 1–5.

From the necessity proof, with the exception of  $m_1$  and  $m_2^{2,*}$ , the strategies in  $s^*$  are identical to the strategies in a SPNE of a game with any  $\epsilon$ . Thus, all we need to do is show that point 6 implies convergence of  $m_1^\epsilon$  to  $m_1^*$  and that point 4 implies convergence of  $m_2^{2,\epsilon}$  to  $m_2^{2,*}$ .

First consider  $m_2^2$ . Fix a  $\delta > 0$ . Now we must show that there exists an  $\epsilon$  such that

$|m_2^{2,\epsilon}(m_1, a_1) - m_2^{2,*}(m_1, a_1)| < \delta$ , for all  $(m_1, a_1)$ . By point 4  $m_2^{2,*}(m_1, a_1) = \hat{m}_2^2(a_1)$  for all  $(m_1, a_1)$ . Thus, it suffices to show there exists an  $\epsilon$  such that  $|m_2^{2,\epsilon}(m_1, a_1) - \hat{m}_2^2(a_1)| < \delta$ .

Consider an  $\epsilon$  and an  $(m_1, a_1)$  such that  $m_2^{2,\epsilon}(m_1, a_1) = m_1^2 \neq \hat{m}_2^2(a_1)$ . The principal's expected utility is  $U_P(a_1^*(m_1^1), a_2^*(m_1, (m_1^1, m_1^2), a_1^*(m_1)))$  which is less than  $U_P(a_1^*(m_1), a_2^*(m_1, (m_1^1, \hat{m}_2^2(a_1)), a_1^*(m_1)))$ . Call the difference  $\gamma$ . Now select a new  $\epsilon' < \gamma$ , and  $m_2^{2,\epsilon'}(m_1, a_1) = \hat{m}_2^2(a_1)$ .

The fact that  $m_1$  converges now follows immediately from the fact that  $a_1^\epsilon$  converges to  $a_1^*$ ,  $a_2^\epsilon$  converges to  $a_2^*$  and  $m_2^\epsilon$  converges to  $m_2^*$ . ■

**Lemma C.5** *Consider a game in which  $a_1$  and  $a_2$  are chosen simultaneously in a single policy-making stage, there is one agent, and observable actions. An associated SPNE,  $s^* = (a_1^*(\cdot), a_2^*(\cdot), m_1^*(\cdot, \cdot, \cdot), \rho^*(\cdot, \cdot, \cdot))$  is robust to small revision costs if and only if*

(i)  $\rho^*(m_1, a_1, a_2, O)(\cdot, \cdot, \cdot) = m_1(\cdot, \cdot, \cdot)$  for all  $(m_1, a_1, a_2, O)$ ;

(ii)  $(a_1^*(m_1), a_2^*(m_1)) \in \arg \max_{(a_1, a_2)} \mathbb{E}[m_1(a_1, a_2, O)\bar{B} + u(\bar{a} - a_1 - a_2)]$  for all  $m_1$ ; and

(iii)  $m_1^*(\cdot, \cdot, \cdot) \in \arg \max_{r \in \mathcal{R}} p(a_1^*(r), a_2^*(r))$ .

**Proof.**

**Necessity:** Fix an  $s^*$  that is robust to small revision costs.

Consider the retention stage. Fix an  $\epsilon$ . At this stage, the agent's actions are already taken. Hence, the principal's action at this stage has no effect on outcomes. If the principal chooses  $\rho^\epsilon(m_1, a_1, a_2, O) \neq m_1$ , she suffers a cost  $\epsilon$ . Hence, sequential rationality requires that  $\rho^\epsilon(m_1, a_1, a_2, O) = m_1$  for all  $(m_1, a_1, a_2, O)$ . Since this is true for all  $\epsilon$  and the sequence converges by hypothesis, this establishes point 1.

Next consider the agent's choice of actions. For any  $\epsilon$ , the agent infers that the principal will choose  $\rho^\epsilon(m_1, a_1, a_2, O) = m_1$ . Hence, the agent chooses  $(a_1^\epsilon(\cdot), a_2^\epsilon(\cdot))$  as a best response to  $m_1$ . Since, given an  $m_1$ , the set of agent best responses is unaffected by the presence of the retention cost  $\epsilon$ , this implies  $(a_1^\epsilon(\cdot), a_2^\epsilon(\cdot)) \in \arg \max_{(a_1, a_2)} \mathbb{E}[m_1(a_1, a_2, O)\bar{B} + u(\bar{a} - a_1 - a_2)]$  for all  $m_1$ . Since  $\rho^\epsilon(m_1, a_1, a_2, O) = \rho^*(m_1, a_1, a_2, O) = m_1$  for all  $(m_1, a_1, a_2, O)$  this implies that  $(a_1^\epsilon(m_1), a_2^\epsilon(m_1)) = (a_1^*(m_1), a_2^*(m_1))$  for all  $m_1$ . Since this is true for all  $\epsilon$  and the sequence converges by hypothesis, this establishes point 2.

Finally, consider the principal's play in the first policy-making stage. Fix an  $\epsilon$ . For any  $m_1$ , the agent chooses  $(a_1^\epsilon(m_1), a_2^\epsilon(m_1)) \in \arg \max_{(a_1, a_2)} \mathbb{E}[m_1(a_1, a_2, O)\bar{B} + u(\bar{a} - a_1 - a_2)]$ . Suppose  $m_1^\epsilon$  violates the condition in point 3. This implies that the principal would be strictly better off choosing a different  $m_1$ . Hence, it must be that, for all  $\epsilon$ ,  $m_1^\epsilon$  satisfies point 3.

**Sufficiency:**

Fix an  $s^*$  satisfying points 1–3. From the necessity proof, we know that the strategies in  $s^*$  are identical to the strategies in a SPNE of a game with any  $\epsilon$ . Hence, if  $s^*$  satisfies points 1–3 it is trivially the limit of a sequence of SPNE in games with ever small revision costs. ■