Policy Making with Reputation Concerns

Qiang Fu†
National University of Singapore

Ming Li‡
Concordia University and CIREQ

First draft: July 2008
This version: July 2010

Abstract

We study the policy choice of an office-holding politician who is concerned with the public's perception of his capabilities. The politician decides whether to maintain the status quo or to conduct a risky reform. The success of the reform depends critically upon the abilities of the politician, which is privately known to the politician. The public observes both his policy choice and the outcome of the reform, and forms a posterior on the true ability of the politician. We show that politicians may engage in socially detrimental reform in order to be perceived as more capable. Conservative institutions that thwart reform may potentially improve social welfare.

JEL Nos: C72, D72, D82
Keywords: Reform, Reputation, Ability, Conservatism

*We are grateful to Parimal Bag, Michael Baye, Oliver Board, Kim-Sau Chung, Rick Harbaugh, Navin Kartik, Kyungmin Kim, Tilman Khumpp, Jingfeng Lu, Marco Ottaviani, Ivan Png, Larry Samuelson, Joel Sobel, and Bernard Yeung for their helpful comments. We gratefully acknowledge valuable input from the seminar and conference audiences at the Midwest Economic Theory Meeting–Fall 2008, the Canadian Economic Association Meeting–2009, the Far Eastern Meeting of the Econometric Society–2009, the Montreal Theory Reading Group, the Hong Kong University of Science and Technology, and Stony Brook Workshop on Political Economy–2010. Part of the research was completed while Fu was a CIREQ visitor, and he is grateful for the hospitality.

†Department of Business Policy, National University of Singapore, 1 Business Link, SINGAPORE, 117592, Tel: (65)65163775, Email: bizfq@nus.edu.sg.

‡Department of Economics, Concordia University, 1455 Boulevard de Maisonneuve Ouest, Montreal, QC, H3G 1M8, CANADA, Tel: +1(514)8482424 Ext. 3922, Email: mingli@alcor.concordia.ca.
1 Introduction

She (Emma) was not much deceived as to her own skill either as an artist or a musician, but she was not unwilling to have others deceived, or sorry to know her reputation for accomplishment often higher than it deserved.

*Emma*, vol. 1, ch. 6, by Jane Austen, English Author

Love of fame brings about eccentricity, and being eccentric brings danger to oneself; therefore the sages exhorted against the love of fame.

*Xing xin za yan*, Li Bangxian, Chinese Poet

We are often concerned about inferences that people draw about us from the actions we take and/or their consequence. These inferences shape our reputation, and often determines our prospect of success, professional or otherwise. Reputation concerns loom large, perhaps more conspicuously, in the public sector or in non-profit organizations, where formal contracts based on explicit performance-based incentives are usually rare. When people in the position to make public policies act out of their concerns for reputation, however, the public can often get worse off because of it. In this paper, we demonstrate that it is possible that policy makers take on risky policy initiatives (“reforms”) to convince the public of their capability. Further, to prevent potential negative welfare consequences of such risk-taking, it is often necessary to enact institutional conservatism, which thwarts even reforms that are beneficial to the public.

There are many examples that demonstrate the strength and prevalence of reputation concerns. The reputation of a technocrat’s professional competence often determines his ability to either reach a higher rung on the hierarchical ladder, or resume his career in the private sector after his term of service is over.¹ Career politics provide more salient examples of this. The prospect of a politician being re-elected depends to a large extent on the public’s perception of his capabilities. In the aftermath of the economic turmoil, commentators deemed Gordon Brown to have lost his “reputation for economic competence” “through a combination of appallingly bad luck and even worse misjudgment,”² which would eventually cost him his premiership. Alternatively, a politician in office may simply be concerned about how the public evaluates his legacy when he steps down. The concerns on reputation can

¹A bureaucrat in the Securities and Exchange Commission, for instance, often seeks a lucrative job offer from private financial firms after he leaves office.

form one important dimension of the informal incentives, and often motivate economic agents to carry out various activities to enhance their reputation.\textsuperscript{3}

In the current paper, we identify one particular context in which reputation concerns affect the behaviour of policy makers – they may take on risky or innovative initiatives whose success depends on their capabilities, so as to manipulate the public’s perception of their talents. They would do so even though they know that they have low capabilities and hence a poor chance of success. We then investigate institutional remedies for the welfare consequences of such behaviour. As Tereza Capelos (2005) states, “political actors often engage in controversial activities that challenge their reputation.” She points out that politicians risk losing their support “after showing inexperience, or wrong judgment.” We predict that undertaking risky actions can be interpreted as rational attempts on the part of a politician to enhance his reputation. Reputation concerns could cause excessive and inefficient risk taking. In addition to delineating the behaviours under reputation concerns, our analysis further investigate the proper institutional remedy to balance the gain and cost when policy makers are subject to such incentives.

Throughout this paper, the decision maker is generically referred to as a “politician”, who must choose between adopting a risky policy option, referred to as “reform”, and maintaining the status quo.\textsuperscript{4} The performance of reform depends on not only the intrinsic value of the available proposal, but also on the quality of implementation, which is in turn determined by the politician’s capability. For instance, if the U.S. President pushes through a fiscal stimulus plan, which may help rescue the economy out of a recession, the ultimate success of the plan depends on how funds are allocated to optimize its effectiveness.\textsuperscript{5} The politician’s capability level can be either high or low, and is privately known to only himself. His performance is independent of his capability if the status quo is maintained. The performance, however, would reflect his capability when the status quo is abandoned: a new policy increases uncertainty, and makes success dependent on his ability to take appropriate action under each contingency.\textsuperscript{6} A capable politician is better at implementing reform, and therefore more

\textsuperscript{3}For instance, Frederick Sheehan (2009) comments that Alan Greenspan deliberately built up his own reputation of competence in designing monetary policy, and went to great lengths to protect it.

\textsuperscript{4}However, the analysis encompasses a variety of environments, such as a judge who has to decide whether or not to exercise his power to strike down a law, a prosecutor who has to decide whether or not to file charges against a crime suspect, a CEO who has to decide whether or not to implement an expansion plan, or even a doctoral candidate who must decide whether or not to pursue a cutting-edge research project.

\textsuperscript{5}In another example, although the acquisition of Compaq by Hewlett-Packard has proven its merit over the years, it is a widely held belief that the initial fiasco was due to the flawed management of the merger by its CEO at the time, Carly Fiorina.

\textsuperscript{6}We do not model how the politician gathers information. However, a politician’s capability to elicit information from various sources is widely viewed as an important part of leadership. The US presidential historian, Erwin C. Hargrove (1966, pp 70-73 and pp 114-116), paints two completely different pictures
likely to bring success. The public observes both the policy choice of the politician and the resultant performance. It forms an assessment of the politician’s capability based on the two pieces of information. The politician chooses a policy alternative to maximize the public’s perception of his competence.

Our analysis proceeds in two layers. We first depict the equilibrium behaviour of the politician, which is summarized as follows.

- **No-full Separation.** There exists a unique hybrid equilibrium. A high-type politician is always “eager” to reveal more information by undertaking reform: he reforms with probability one whenever a sufficiently valuable proposal is available. The low type, however, mimics his high-type counterpart with a positive probability. Although the reform undertaken by a low type fails with a higher probability, his reputation concerns “force” him to risk, because he would otherwise suffer from a more unfavourable assessment.

- **Pressure to prove oneself.** The low type reforms less frequently when the public holds a more favourable prior on the type of the politician, i.e., when they believe the politician in office is more likely to be capable. Due to this effect, reform can be predicted to occur less often when the initial assessment of the politician is more favourable, without knowing his true type. We interpret these results as an indication of the pressure to prove oneself. This phenomenon is present in the intellectual, political, and social aspects of our lives. We discuss it more extensively in Section 3 along with illustrative examples.

- **Tough act to follow.** The higher the capability differential between the high type and the low type, the less likely is the low type to undertake reform. To put it simply, the widened gap makes successful mimicry more difficult. This effect causes the low type to reform less often in order to avoid failure.

Based on the equilibrium results, we evaluate the welfare remifications of policy making under career concerns. We consider the design of the optimal (welfare-maximizing) institution or bureaucratic rule that restricts the discretion of the politician. The results are summarized as follows.

---

7This assumption can be related to the concept of “state capacity,” which is proposed by Theda Skocpol (1985). She argues that ambitious reform attempts often fail because bureaucrats usually lack the required competence to administer their reform.
• **Thwarted good reforms.** Assume that a legislature enforces a threshold rule – it prohibits reform unless the intrinsic value of the reform proposal exceeds a threshold. A higher, or more conservative, threshold has two competing effects. It discourages an incapable politician from undertaking detrimental reform on one hand; while it prevents a capable politician from undertaking beneficial reform, thereby sacrificing the gain from his efficient innovation on the other. The social optimum requires proper “institutional conservatism”: the optimal threshold rule must thwart otherwise beneficial reform. Our analysis on the optimal institution lends support to the bureaucratic rules in various organizations that restrict the discretionary power of politicians or bureaucrats in carrying out risky activities. It also provides an alternative rationale for the often observed organizational resistance to policy reform and the widely discussed bias towards the status quo. As pointed out by Raquel Fernandez and Dani Rodrik (1991), “one of the fundamental questions in political economy” has been why governments often fail to carry out efficiency-enhancing reform.

• **Opportunities hurt and “optimism” requires more conservatism.** In an environment in which good reform proposals are more likely to emerge, it is not necessarily true that social welfare will be higher. On the one hand, society gains more from the efficient reform undertaken by the high-type. On the other hand, it “forces” the low type to reform more often, as the choice to forego reform will be more likely to be attributed by the public to the politician’s lack of ability, instead of to the lack of opportunities (i.e. the reform proposal is of low value). The joint effect may be that a more favourable environment paradoxically leads to a decrease in social welfare. To remedy this problem, a more conservative bureaucratic rule can be necessary in response to a more favourable environment.

In the rest of this section, we discuss the link between our paper and the relevant literature. In Section 2, we set up the model. In Section 3, we characterize equilibria of the model and present comparative statics of relevant environmental factors. In Section 4, we discuss the welfare implications of our equilibrium results and consider the issue of institution design. In Section 5, we investigate robustness of our findings and compare our paper with closely related work. In Section 6, we conclude. All proofs are collated in the Appendix.

**Relationship to the Literature**

The notion of career or reputation concerns is featured prominently in the pathbreaking work of Bengt Holmström (1982, 1999). Since then, an enormous amount of scholarly ef-

---

For instance, in the debate on judicial restraint or judicial activism, one popular argument is to encourage judges to refrain from exercising their power to strike down existing laws.
fort has been devoted to exploring the incentive effects of reputation or career concerns in a wide array of environments, including corporate decision making (e.g. Bengt Holmström and Joan E. Ricart i Costa 1986, Jeffery Zwiebel 1995, and Adam Brandenburger and Ben Polak 1996), economic agents’ effort supply (e.g. Holmström 1999 and Alberto Alesina and Guido Tabellini 2007), and experts’ strategic advising activities (e.g. Stephen Morris 2001 and Marco Ottaviani and Peter Norman Sørensen 2006). The literature reveals in various contexts that concerns regarding public or market perceptions distort economic agents’ decision making. Such incentives lead economic agents to ignore their own useful information and instead, to strategically manipulate the belief of the public or the market.9

Our paper explores (1) the politician’s incentives to conduct reform, so as to signal his competence; and (2) the optimal (welfare-maximizing) bureaucratic rule that restricts the politician’s discretion when he is subject to reputation concerns. Hence, it belongs to the strand of career concerns literature that focuses on agents’ incentives to undertake a risky project. The setup of our paper is a variation of the example introduced in Section 3.2 of Holmström’s (1999) seminal paper. The common feature is that the politician’s (decision maker’s) talent is only relevant when the reform (risky project) is undertaken. Hence, more information can be revealed when the risky activity is carried out.10 Two features distinguish our setup from Holmström’s (1999): first, we assume the politician’s talent is his private information, while he assumes symmetric information and symmetric information updating; second, in our model, the probability of success for each type is common knowledge, while in Holmström’s (1999), it is the private information of the agent. As a consequence, in our model, the choice to undertake reform can signal the type of the politician, which is not possible in Holmström’s (1999) setup.

Holmström and Ricart i Costa (1986) study managers’ incentives to make a new investment when the manager is subject to career concerns. Benjamin E. Hermalin (1993) shows that a risk-averse agent with career concerns may have the incentive to choose a more risky project. However, in his model, a risky project is a worse indicator of the agent’s talent, while in our model, reform reveals more information. Gary Biglaiser and Claudio Mezzetti (1997) study politicians’ incentives to implement major new projects. Voters evaluate the incumbent’s ability based on his performance in the project. They show that the incumbents’ project choices can be either too radical or too conservative, depending on the bias of the median voters. Similar to Holmström’s (1999), these studies mainly focus on complete-information settings, where the type of the manager is unknown to all players.

9For instance, Brandenburger and Polak (1996), David S. Scharfstein and Jeremy C. Stein (1990), and Ottaviani and Sørensen (2006) all share this feature.

10The inclusion of a “status quo” option that does not reveal the right action to take for the risky option is also present in Amal Sanyal and Kunal Sengupta (2006). They study a game of strategic communication in which the expert is career-concerned in the sense of Ottaviani and Sørensen (2006).
Biglaiser and Mezzetti (1997) also briefly discuss an extension of asymmetric information. They demonstrate the impossibility of full separation but do not fully characterize all the equilibria in that case. Within this strand of literature, our paper is closely linked to Jeffery Zwiebel (1995). He also explores how reputation concerns moderate a manager’s incentive to undertake risky innovation. He shows that managers with intermediate capabilities may resist innovation that can be beneficial to the firm. In his setting, the manager’s innovative action is unobservable. Hence, this context (1995) does not involve signalling action on the part of the decision maker. We also arrive at the opposite conclusion that there can be too much reform.\footnote{Robert A. J. Dur (2001) and Peter Howitt and Ronald Wintrobe (1995) also explore scenarios in which there is too little change in policy.}

Our study includes flavours from both the literature of signalling and that of career concerns, which places it in the company of a handful of other studies. They include the notable examples of Canice Prendergast and Lars Stole (1996), Gilat Levy (2007), and Wei Li (2007).\footnote{Prendergast and Stole (1996) argue that career concerns induce young investors to overreact to new information they receive, so as to signal that they are fast learners. Wei Li (2007) makes a similar point in the case of experts providing advice to decision makers. We discuss Levy’s work at the end of the literature discussion.} In a recent paper, Kim-Sau Chung and Péter Esö (2008) build a model in which a worker chooses a task to both signal his capabilities to potential employers and learn about his capabilities himself, as he has only imperfect knowledge of it. They assume that the more difficult task is a worse (less informative) device for assessing the capability of a worker; meanwhile in our setting, undertaking the more difficult task (reform) allows for more information transmission.

Sumon Majumdar and Sharun W. Mukand (2004) and Guido Suurmond, Otto H. Swank, and Bauke Visser (2004) both consider the incentives of agents in the public sector to undertake risky projects, which signal their types. Majumdar and Mukand (2004) study the dynamic incentives of a government within an election cycle and its policy persistence. The government can be either too radical or too conservative in equilibrium. Suurmond, Swank, and Visser (2004) contend that the presence of career concerns can be socially beneficial, as it can encourage a smart agent to expend more effort in gathering information. Ying Chen (2010), in a simultaneous and independent paper, analyzes the choice of an agent chooses between a risky project and a safe project. Chen (2010) also assumes that the likelihood of success of the risky project depends on the agent’s ability. Our paper and Chen’s (2010) are different from each other in a few respects. Chen (2010) mainly focuses on the impact of information structure on the agent’s project choice under career concerns. She shows that the agent takes excessive/inadequate risks when he does/does not know his own type. In contrast, we focus on a setting in which the politician knows his type. In addition
to identifying the problem of excessive risk taking, we focus more on the roles of various environmental factors in determining the politician’s behaviour and the design of optimal institution that remedies this problem. Besides the difference in focuses, our study exhibits different modeling characteristics from Majumdar and Mukand (2004), Suurmond, Swank, and Visser (2004) and Chen (2010). The modeling difference and the roles played the unique flavours of our model will be discussed in Sections 2 and 5.

Our analysis of optimal institution design in the presence of reputation concerns is conceptually related to that in a small number of other papers, which study the ramifications of various institutions in career-concerns models. Andrea Prat (2005) argues that transparency in an organization may hurt as the agent may take revealed action to influence the principal’s posterior instead of seeking the best interests of the organization. Gilat Levy (2007) shows that in a committee of voters with career concerns, radical actions are more likely to be accepted when the voting process is transparent. To our knowledge, our paper might be one of the first to explicitly investigate an institutional remedy for inefficient risk taking when the decision maker has reputation concerns. Felix Bierbrauer and Lydia Mechtenberg (2008) analyze the welfare effect of early elections when the political leaders have career concerns. Our result that restrictions on changes to the status quo could be welfare-improving complements other justifications of institutional conservatism, for example, those offered by Li, Hao (2001) and Young K. Kwon (2005). Our analysis suggests that institutional barriers (bureaucracy) that limit the amount of discretion that the decision maker can exercise may be welfare-enhancing. Our paper echoes the conclusion of Jean Tirole (1986) in this respect.

2 Setup

A politician makes a policy choice between two alternatives: maintaining the status quo or initiating a reform. If the politician retains the status quo, the outcome of this polity, $y$, is deterministic, which we normalize to 0. In contrast, if the politician chooses to undertake the reform, uncertainty will arise and the politician must take an action to address it. The outcome of a reform is given by quadratic loss function

$$y = \theta - (a - \omega)^2.$$  

(1)

where $\theta$ measures the intrinsic value of the available reform proposal, $\omega$ is the true state of the world, and $a$ is the action taken by the politician in response to his assessment of $\omega$.

The intrinsic value of the available reform proposal, $\theta$, is continuously distributed on $[-\theta_1, \theta_2]$ with a distribution function $F$ and density function $f$, where $-\theta_1 < 0 < \theta_2$ and $\theta_1, \theta_2 \in (1, 2)$. The distribution of $\theta$ is common knowledge. The realization of $\theta$ is observed by the politician before he decides whether or not to adopt the reform proposal, while it is
unobservable or *ex ante* unverifiable to the general public.

The ultimate performance of the reform depends on not only the intrinsic value of the reform proposal, but also the quality of politician’s implementation, i.e., how well he addresses the uncertainty that arises with reform. The uncertainty is embodied by the state of the world, \( \omega \), which may take either of two values from \( \Omega = \{-1, 1\} \), each with a probability \( \frac{1}{2} \). The state \( \omega \) is realized only after a reform has been initiated. The politician has to choose his action \( a \) from \( A = \{-1, 1\} \) to implement the reform. We say that the reform is a *success* if his action matches the state of the world, while it is a *failure* if it does not. Neither the politician nor the public observes the true state. However, the politician can receive a signal \( \sigma \in \{-1, 1\} \) about \( \omega \). The precision of the signal depends on the talent of the politician.

The talent of the politician, \( t \), is drawn from the set \( \{L, H\} \). A high-talent politician \( (H) \) receives an informative signal, which matches the true state with a probability \( q = \Pr(\sigma = \omega) > \frac{3}{4} \). In contrast, a low-talent politician’s signal is completely uninformative. It should be noted that the assumption \( q > \frac{3}{4} \) does not affect our analysis. However, without this assumption, no reform can be socially beneficial. The talent of the politician is his private information. Let \( \alpha \) be the probability of \( t = H \), which is commonly known. It is the public’s prior about the politician’s talent, which can also be viewed as the proportion of high-capability politicians in the “population.”

\[ \alpha < \frac{1}{2}. \]

Upon receiving \( \sigma \) (either informative or uninformative), the politician takes an action.

The public observes the politician’s policy choice (status quo or reform) and the final outcome \( y \). By the assumptions of \( -\theta_1 < 0 < \theta_2 \) and quadratic output function, \( \theta \) can be ex post inferred from \( y \) if and only if the reform is carried out. The public’s posterior or the reputation of the politician, can be written as

\[ \bar{\mu}^i(y) \equiv \Pr(t = H|y, i) \]

---

13 There is literature that analyzes the composition of politicians as a group, which is complementary to our research, in that it offers an explanation for why politicians may consist of a significant proportion of low-ability individuals. Francesco Caselli and Massimo Morelli (2004), Matthias Messner and Mattias K. Polborn (2004), and Andrea Mattotzzi and Antonio Merlo (2007, 2008) have offered various explanations for why political processes tend to select low-ability individuals to be politicians.

14 This regularity assumption is only required so that in the extreme case where the high type’s signal is perfectly informative, the low type still has an incentive to undertake reform and mimic the high type (see the proof of Part 1 of Proposition 1.)

15 In our setup, whether or not the public observe the action is inconsequential. Once the politician chooses reform, the belief of the public is determined only by whether the outcome is a “failure” or “success.”
by Baye’s rule, where \( i = 0 \) indicates the status quo and \( i = 1 \) indicates reform. Borrowing from much of the career-concerns literature, we assume that the politician’s payoff depends purely on his reputation.

We assume that the politician has only limited discretion. Specifically, he may adopt a reform proposal only if its value exceeds a cutoff \( \hat{\theta} \in [-\theta_1, \theta_2] \). We implicitly assume that the politician’s policy choice is subject to the regulation or monitoring of a “legislature” (e.g., parliament, supreme court, an advisory committee, a board of directors). The legislature cannot verify the type of the politician, but it abides by certain institutional rule that constrains the politician’s authority or discretion. The rule can be understood as a constitution, or as widely observed, an organizational bureaucracy (see Tirole 1986), which prevents him from undertaking apparently harmful activities. Such institutional restrictions are prevalent in various organizations. For instance, the US President must obtain congressional approval for his policy choices. Military commanders have to honour “rules of engagement” in the use of force. An administrator of the Environmental Protection Agency has to rely on limited authority and resources to regulate polluting industries. Finally, judges are often pressured to refrain from exercising their power to strike down existing laws.

Under the institutional constraint (e.g., constitution), the politician is subject to the “rule of engagement” and is authorized to undertake a reform only if the intrinsic value of the available reform proposal exceeds the cutoff \( \hat{\theta} \). Analogous to Tirole (1986), it is implicitly assumed that the politician must provide a (partially verifiable) report on the quality (\( \theta \)) of his reform proposal to the legislature when he pushes forward a reform, although such information is neither verifiable nor accessible ex ante to the general public. To abide by the “constitution”, the legislature would not approve any reform proposal with a value below \( \hat{\theta} \). For the moment, we assume that \( \hat{\theta} \) is fixed and focus on the equilibrium behaviour of the politician. We dedicate Section 4 to an in-depth analysis on the welfare-maximizing rule \( \hat{\theta}^* \), which endogenizes the cutoff.

The timeline of the model is as follows. First, nature chooses the quality of the reform proposal, i.e., the value of \( \theta \). Second, the politician observes \( \theta \) and decides whether to adopt the reform proposal. He further chooses \( a \) if he decides to undertake the reform. Third, the public updates their belief after observing both the politician’s policy choice and performance.

The setup of our model differs from the existing literature in a number of ways. Here, we stress two essential features of our model. First, the distinction between policies (status quo or reform) and actions is important in our model. Policies are macro-level or “strategic” decisions such as whether to reform financial regulations or whether to start a war. In contrast, actions are micro-level or “tactical” decisions such as which instrument of regulation to introduce in overhauling the financial system or how many troops to deploy in the war.
The true nature of the problem \((\omega)\) determines which action is ex post suitable for implementing the reform. Second, in contrast to many existing career-concerns models with risky experimentation (e.g. Majumdar and Mukand 2004; Suurmond, Swank, and Visser 2004), the outcome of a reform is measured by a continuous variable, instead of a binary indicator (e.g., success or failure) alone. It depends on both the quality of the reform proposal \((\theta)\) and the quality of implementation \((|a - \omega|)\), while both are subject to random perturbation. This setup enriches our analysis in two aspects. First, it enables an analysis of institution design. A more sophisticated trade-off is involved in determining the proper level of institutional conservatism. Second, a comparative static analysis may be performed on the probability distribution of the value of reform, which sheds further light on the equilibrium behaviour and the design of welfare-maximizing institution.

3 Equilibrium Analysis

In this part, we first study the benchmark of the first best situation in which the public’s expected payoff from the politician’s policy choice is maximized. We then derive the equilibrium and conduct comparative analysis.

3.1 First Best Benchmark

Let us define

\[
q_t = \begin{cases} 
q & \text{for } t = H; \\
\frac{1}{2} & \text{for } t = L.
\end{cases}
\]

When the politician chooses to reform and takes an action \(a\), the expected outcome of the reform is given by

\[
E(y) = \theta - E_{\omega \in \{-1, 1\}}(a - \omega)^2
= \theta - 4(1 - q_t)
\]

A high-type politician maximizes his probability of success by following his signal, i.e., choosing \(a = \sigma\). A low-type politician’s signal is uninformative and the two states are equally likely. His choice of \(a\) is \textit{ex ante} irrelevant.

In the first-best situation, a politician would undertake reform if and only if the expected outcome \(E(y)\) is non-negative. A low-type politician should never reform regardless of \(\theta\) as the expected loss from wrong actions always exceeds the benefit of reform, that is,

\[
E(y) = \frac{1}{2}\theta + \frac{1}{2}(\theta - 4) = \theta - 2 < 0,
\]
because $\theta \leq \theta_2 < 2$. The expected outcome for a high-type politician is given by

$$E(y) = \theta - 4(1 - q).$$

The high type should undertake reform if and only if the value of reform is sufficiently high, i.e., $\theta \geq 4(1 - q)$.

### 3.2 Equilibrium

We adopt the concept of *Divine Equilibrium*, first introduced by Jeffrey S. Banks and Joel Sobel (1987). The equilibrium requires (1) the politician and the public form Bayesian beliefs, (2) the politician chooses the action that maximizes his expected reputation if he undertakes reform, (3) the politician chooses reform or the status quo to maximize his expected reputation; and (4) the out-of-equilibrium belief satisfies “divinity” criterion. “Divinity” criterion imposes mild and sensible restrictions on out-of-equilibrium beliefs, which distinguishes itself from a Perfect Bayesian Equilibrium. However, “divinity” criterion is weaker than the D1 criterion of Banks and Sobel (1987). Further detail is provided in the Appendix.

#### 3.2.1 Equilibrium

We consider a politician’s behavioral strategy. Let $\rho_t(\theta)$ be the probability with which a type-$t$ politician chooses reform when its value is $\theta$. As implied by the institutional rule, $\rho_t(\theta) = 0$ for $\theta \in [-\theta_1, \hat{\theta})$, for $t \in \{L, H\}$. When the politician maintains the status quo, his reputation among the public is

$$\mu^0 = \frac{\alpha F(\hat{\theta}) + \alpha \int_{\theta_2}^{\hat{\theta}} [1 - \rho_H(\theta)] f(\theta) d\theta}{F(\hat{\theta}) + \alpha \int_{\theta_2}^{\hat{\theta}} [1 - \rho_H(\theta)] f(\theta) d\theta + (1 - \alpha) \int_{\theta_2}^{\hat{\theta}} [1 - \rho_L(\theta)] f(\theta) d\theta}.$$

Note that, as long as reform is undertaken, the public can *ex post* perfectly infer the value of $\theta$ from the outcome $y$. With a slight abuse of notation, when the politician implements a reform of value $\theta$, his reputation will become

$$\mu^s = \frac{\alpha q \rho_H(\theta) f(\theta)}{\alpha q \rho_H(\theta) f(\theta) + (1 - \alpha) \frac{1}{2} \rho_L(\theta) f(\theta)}$$

if the reform succeeds, and

$$\mu^f = \frac{\alpha (1 - q) \rho_H(\theta) f(\theta)}{\alpha (1 - q) \rho_H(\theta) f(\theta) + (1 - \alpha) \frac{1}{2} \rho_L(\theta) f(\theta)}$$

if the reform fails. If a type-$t$ politician implements a reform with value $\theta$, he receives an expected payoff

$$\mu_t = q_t \mu^s + (1 - q_t) \mu^f.$$

The following proposition characterizes the full set of equilibria.
Proposition 1. 1. There exists a unique equilibrium of the game. In the equilibrium, the high-type politician undertakes reform with probability $\rho^*_H(\theta) = 1$ whenever he receives a proposal of value $\theta \in [\hat{\theta}, \theta_2]$ and $\rho^*_H(\theta) = 0$ otherwise, while the low-type politician undertakes reform with probability $\rho^*_L(\theta) = \rho^* \in (0, 1)$ when $\theta \in [\hat{\theta}, \theta_2]$ and $\rho^*_L(\theta) = 0$ otherwise.

2. The equilibrium probability $\rho^*$ solves

$$\frac{1}{1 + \lambda(\alpha)A} = \frac{1}{2} \cdot \frac{1}{1 + \lambda(\alpha)B} + \frac{1}{2} \cdot \frac{1}{1 + \lambda(\alpha)C^*},$$

where

$$\lambda(\alpha) = \frac{1 - \alpha}{\alpha}, \ A = 1 + (1 - \rho)\kappa(\hat{\theta}), \ \kappa(\hat{\theta}) = \frac{1 - F(\hat{\theta})}{F(\hat{\theta})}, \ B = \frac{\frac{1}{2}\rho}{q}, \ C = \frac{\frac{1}{2}\rho}{1 - q}.$$

Proposition 1 states that there can be no full separation of the two types. The high type is always “eager” to undertake reform: he does so whenever a favorable proposal is received, i.e. $\theta \geq \hat{\theta}$. The low type always mimics his high-type counterpart and undertakes reform with a positive probability, $\rho^*$.

The policy choice of the low type involves subtle trade-offs. Neither fully pooling nor fully separating is possible in equilibrium. Despite the low type’s incompetence, he still hopes for accidental success, and his refusal to take reform is likely to be read as a signal of incompetence. His concern for reputation thus compels him to take the risk. However, failure to take reform would not fully reveal his type because the true realization of $\theta$ is not observable to the public if reform is not taken and $\theta$ can fall below $\hat{\theta}$. He thus randomizes in the equilibrium.

The nature of the strategic concerns is better revealed when we make comparison across equilibria. With a slight abuse of notation, let us denote by $E_{\mu_t}(\hat{\theta})$ the ex ante expected payoff of a type-$t$ politician in an equilibrium with a given $\hat{\theta}$. Our analysis leads to the following proposition.

Proposition 2. 1. The equilibrium probability of reform by the low type, $\rho^*$, strictly decreases with $\hat{\theta}$.

2. The low-type politician always prefers an equilibrium with a higher cutoff $\hat{\theta}$, while the high-type politician always prefers an equilibrium with a lower $\hat{\theta}$. That is, $\frac{dE_{\mu_H}(\hat{\theta})}{d\hat{\theta}} < 0$, and $\frac{dE_{\mu_L}(\hat{\theta})}{d\hat{\theta}} > 0$.

The politician faces a more stringent standard for taking reform when a higher $\hat{\theta}$ is present. Overall, the high type reforms less often when $\hat{\theta}$ increases. Hence, the low type obtains higher reputation from maintaining the status quo. A no-reform outcome is more
likely to be interpreted by the public as the result of no opportunity for reform ($\theta < \bar{\theta}$) rather than low capability, which causes him to reform with a smaller probability.

Part 2 of Proposition 2 describes the two types’ ranking of equilibria under different cutoffs. The high type prefers an equilibrium with a lower cutoff, where he can reform more, because it ensures that his capability is revealed with a higher probability. The low type, however, prefers an equilibrium with a higher cutoff hence less reform, due to two reasons. First, lower frequency of reform allows the low type to pool with his high-type counterpart more often and to reveal less information. Second, when the low type reforms less often, the public would believe that a reform is increasingly likely to be implemented by the high type, which mitigates the damage to the low type when the reform fails.

3.2.2 Comparative Statics

We now examine how the politician’s equilibrium behaviour varies with environment parameters. In the equilibrium, the high-type politician reforms with probability one whenever $\theta$ exceeds $\hat{\theta}$, while the low type reforms with a probability $\rho^*$. Hence, in this equilibrium, reform occurs with a probability

$$\bar{\rho} = [1 - F(\hat{\theta})][\alpha + (1 - \alpha) \rho^*_{|\theta = \hat{\theta}}]. \hspace{1cm} (4)$$

The main results are summarized in the following proposition.

**Proposition 3.** Consider the equilibrium under a threshold rule $\hat{\theta}$.

1. The probability of reform by the low type, $\rho^*$, is strictly decreasing in $\alpha$, the public’s prior. The overall likelihood of reform, $\bar{\rho}$, also strictly decreases with $\alpha$.

2. The probability of reform by the low type, $\rho^*$, is strictly decreasing in $q$. The overall likelihood of reform, $\bar{\rho}$, also strictly decreases with $q$.

3. Let $\rho$ and $\rho'$ denote, respectively, the equilibrium probabilities of the low type undertaking reform associated with distributions $F(\cdot)$ and $G(\cdot)$. Let $\bar{\rho}$ and $\bar{\rho}'$ be their counterparts for the overall likelihood of reform. For a given $\bar{\theta}$, then, $\rho > \rho'$ and $\bar{\rho} > \bar{\rho}'$ if $F(\cdot)$ first order stochastically dominates $G(\cdot)$.

Now, we discuss the intuition and implications of these results. Part 1 of Proposition 3 states that the low type conducts more reforms, when the public holds a less favourable prior assessment, or the proportion of capable politicians in the population is smaller. A more favourable prior assessment increases a politician’s loss from a failed reform, which consequently weakens his incentive to reform. By contrast, a less favourable prior assessment would strengthen his incentive to take risk, because it implies a smaller loss from a failed
reform but a larger gain from an accidental success. This is then interpreted as the pressure to prove oneself phenomenon.

Part 1 of Proposition 3 further shows that less reform would take place overall when the public has a more favourable assessment of the politician’s talent (or there is a higher proportion of capable politicians). Note that

$$\frac{\partial \bar{\rho}}{\partial \alpha} = (1 - F(\hat{\theta}))[1 - \rho^* + (1 - \alpha)\frac{\partial \rho^*}{\partial \alpha}].$$  \hspace{1cm} (5)

Two competing forces come into play when $\alpha$ is higher. On the one hand, since the low type reforms less often than the high type, more reform would be expected when there is a higher proportion of the high type. This effect is depicted by the term $(1 - \rho^*)$. On the other hand, a larger $\alpha$ leads the low type to reform less, which tends to reduce the frequency of reform. The latter effect is witnessed in the term $(1 - \alpha)\frac{\partial \rho^*}{\partial \alpha}$. Our analysis shows that the latter effect always dominates the former one.

This result yields an empirically testable hypothesis: when there is a smaller proportion of capable politicians in the population or when the public holds a more pessimistic prior view, more reform is expected. Conversely, the public observes less reform when the politician has a better reputation. This conclusion is drawn without knowledge of the true type of the politician, which is his private information and is unverifiable.

This phenomenon can be witnessed in a wide variety of contexts, and our result sheds light on it. Young or less established individuals are usually seen as being more progressive and opposed to the status quo, in contrast to senior or more established individuals, who usually behave more prudently and conservatively. A famous example is the “Young’ Turks” reform movement, which agitated against the Ottoman Empire in the early 20th century, thus building a rich tradition of dissent and paving the foundation of modern Turkey.\footnote{The Young Turks originated from the secret societies of progressive and modernist university students and military cadets, who advocated reformation of the Ottoman administration and promoted social and political changes against the monarchy. The Young Turk revolution re-established the constitutional era in 1908. As a nationalist party, Young Turks dominated the domestic politics of Turkey thereafter for an entire decade.}

The term “Young Turks” today represents progressive individuals who are eager to bring about widespread change.

Our analysis may also account for the controversial and seemingly “imprudent” move of Lee Hsian Loong (the current prime minister of Singapore) in 2004. Lee visited Taiwan and demonstrated conspicuously his interests in mediating Sino-Taiwan relation. This move apparently deviated from the long-lasting policy paradigm set by his father Lee Kuan Yew (the founding father of Singapore), which committed to Singapore’s “non-involvement” with the cross-strait affairs. His visit enraged China, and caused turbulence to the country’s relation to China. Political commentators regarded his visit as being “unnecessary” and
the demonstration of lack of “diplomacy and delicacy” in handling international relations. Commentators have offered various possible explanations of his actions. A rationale, however, can be found in light of the pressure to prove oneself phenomenon. Lee’s visit coincided with the official confirmation of his prime ministership. An analogy was often drawn between his rise and dynastic succession. This move can be plausibly interpreted as an attempt to establish his credibility and independence.

Part 2 of Proposition 3 states that a low-type politician would mimic his high-type counterpart less often when the latter becomes more capable. The logic of this result is as follows. When the high type has a more accurate signal, the public is more likely to attribute an unsuccessful reform to a low-type politician, which unambiguously increases his cost of conducting reform. To put it simply, a higher $q$ makes it more difficult for a low type to mimic his high-type counterpart, and therefore leads to a lower probability of reform by the low type. This result is interpreted as the tough act to follow phenomenon.

The distribution of $\theta$ does not qualitatively alter the main prediction of our analysis, but it quantitatively affects the equilibrium behaviour. Part 3 of Proposition 3 describes its effect on $\rho^*$. A stochastically dominant distribution implies that the probability mass is shifted upward. Hence, favourable reform proposals are more likely to be realized. Given the better prospect of reform, the public would then believe that a no-reform outcome is more likely to be caused by the politician’s lack of talent, instead of a lack of opportunities (a lower realization of $\theta$). The public’s assessment of the politician’s ability is therefore lowered when they observe no reform, and this “forces” the low type to reform more often. This result yields interesting welfare implications, which are discussed later in this paper.

4 Institutional Design

In our analysis the legislature does not use contingent monetary transfer to elicit desirable action. The legislature abides by the “constitution” $\hat{\theta}$. Based on the equilibrium analysis above, we turn to the investigation of the optimal (welfare-maximizing) institution that governs the politician’s scope of discretion.

In our model, a higher $\hat{\theta}$ represents a more conservative rule that grants less authority to the politician; while a lower $\hat{\theta}$ represents a more liberal rule that is more permissive of reform. As aforementioned, the society may expect a gain from the reform that is undertaken by the high-type politician (when $\theta$ is sufficiently high), while it always expects a loss from the reform that is undertaken by the low type. A trade-off is triggered when a more conservative rule is adopted. By restricting reform, it reduces the damage from the latter on the one hand, while it also leads to less gain from the former on the other. Under an arbitrary threshold
rule \( \hat{\theta} \), the social welfare in this equilibrium can be written as a function

\[
W = \alpha \int_{\hat{\theta}}^{\theta_2} [\theta - 4(1 - q)]f(\theta)d\theta + (1 - \alpha)\rho^* \int_{\hat{\theta}}^{\theta_2} (\theta - 2)f(\theta)d\theta.
\]

(6)

By implementing a proposal of value \( \theta \geq \hat{\theta} \), the high-type politician contributes an expected outcome of \( \theta - 4(1 - q) \), while the low-type generates a loss of \( \theta - 2 \). The term \( W_1 \) thus represents the overall net gain from the reform that is undertaken by the high type; while the term \( W_2 \) depicts the overall loss from the inefficient reform that is undertaken by the low type. Clearly, the optimal rule \( \hat{\theta}^* \) must exceed \( 4(1 - q) \).

Consider an arbitrary reform proposal with a value \( \theta \in [\hat{\theta}, \theta_2] \). The ex ante expected outcome of this proposal under the threshold rule \( \hat{\theta} \) is given by

\[
E(y|\theta, \hat{\theta}) = \alpha[\theta - 4(1 - q)] + (1 - \alpha)\rho^*(\theta - 2),
\]

which, for a given \( \hat{\theta} \), strictly increases with \( \theta \). Define \( \rho \equiv \lim_{\theta \uparrow \theta_2} \rho^* \). We have the following.

**Lemma 1.** Whenever

\[
\frac{(1 - \alpha)\rho}{\alpha} < \frac{\theta_2 - 4(1 - q)}{2 - \theta_2},
\]

(7)

there exists a unique \( \hat{\theta}^0 \in (4(1 - q), \theta_2) \) that solves

\[
E(y|\hat{\theta}, \hat{\theta}) = \alpha[\hat{\theta} - 4(1 - q)] + (1 - \alpha)\rho^*(\hat{\theta} - 2) = 0.
\]

Further, \( \hat{\theta}^0 \) exhibits the following property: for any \( \hat{\theta} \in [-\theta_1, \theta_2] \),

\[
E(y|\hat{\theta}, \hat{\theta}) \gtrless 0 \text{ if and only if } \hat{\theta} \gtrless \hat{\theta}^0 .
\]

(8)

Consider an arbitrary threshold rule \( \hat{\theta} \). The expression in (8) depicts the expected outcome from a “marginal” reform proposal, i.e. the proposal with a value of exactly \( \hat{\theta} \). The property of \( \hat{\theta}^0 \) demonstrated by (8) yields interesting implications. Specifically, the threshold rule \( \hat{\theta}^0 \) can be used as a natural benchmark. If the prevailing rule \( \hat{\theta} \) is less conservative than \( \hat{\theta}^0 \), it must admit “bad” reform: reform with a value in \( [\hat{\theta}, \hat{\theta}^0) \) would be allowed, which yields negative expected outcome.\(^{17}\) In contrast, if the prevailing rule \( \hat{\theta} \) imposes more restrictions than \( \hat{\theta}^0 \), it must thwart otherwise “good” reform: reform with a value in \( [\hat{\theta}^0, \hat{\theta}) \) would be prohibited, which would otherwise yield a positive expected outcome. However, a threshold rule \( \hat{\theta}^0 \), by its very definition, can be considered as a “neutral” cutoff: it completely rules out “bad” reform, while it does not thwart otherwise beneficial reform. So, is \( \hat{\theta}^0 \) the optimal

\(^{17}\)Based on the definition of \( \hat{\theta}^0 \), under the threshold \( \hat{\theta} < \hat{\theta}^0 \), even a reform with a value that is higher than \( \hat{\theta}^0 \) may still incur an expected loss.
cutoff $\hat{\theta}^*$ that maximizes social welfare? If not, then would the optimal institution be more conservative or less conservative, i.e., does the optimum require $\hat{\theta}^* < \theta^0$ or $\hat{\theta}^* > \theta^0$?

Our analysis yields the following result.

**Proposition 4.** A unique socially optimal cutoff $\hat{\theta}^* \in (\hat{\theta}^0, \theta_2)$ exists if and only if (7) is satisfied; otherwise, the public prefers no reform at all, i.e., $\hat{\theta}^* = \theta_2$.

This proposition states that a unique optimal threshold exists, and the optimum $\hat{\theta}^*$ must exceed $\hat{\theta}^0$ whenever $\hat{\theta}^0$ exists. The welfare maximizing institutional rule requires more conservatism than $\hat{\theta}^0$. In order to understand its logic, let us now analyze the marginal impact of an increase in $\hat{\theta}$ on social welfare. Taking the first order derivative of (6) with respect to $\hat{\theta}$ yields

$$\frac{dW}{d\hat{\theta}} = f(\hat{\theta}) \left\{ \begin{array}{c} -\alpha[\hat{\theta} - 4(1-q)] - (1-\alpha)\rho^*(\hat{\theta} - 2) \\ + (1-\alpha) \frac{d\rho^*}{d\hat{\theta}} \int_{\theta_1}^{\theta_2} (\theta - 2)f(\theta)d\theta \end{array} \right\}. \quad (9)$$

An increase in $\hat{\theta}$ affects $W$ through three venues. First, it reduces the beneficial reform that is undertaken by the high type, and therefore decreases the gains from reform by the high-type politician. This loss is shown by the term $a$, which is negative whenever $\hat{\theta} > 4(1-q)$. Second, a higher cutoff $\hat{\theta}$ (directly) reduces the expected loss from the inefficient reform that is undertaken by the low type. This (direct) effect is embodied by the term $b$. Third, it leads the low-type politician to refrain from undertaking reform for any given $\theta \geq \hat{\theta}$ (because $d\rho^*/d\hat{\theta} < 0$ by Lemma 2), which further reduces the loss from the inefficient reform that is undertaken by the low type. This positive (indirect) effect is depicted by the term $c$.

The decomposition of $dW/d\hat{\theta}$ demonstrates that $\hat{\theta}^0$ is never the optimal threshold. When $\hat{\theta} = \hat{\theta}^0$ is enforced, social welfare can be increased by raising $\hat{\theta}$: the sum of the first two terms simply boils down to $E(y|\theta^0, \hat{\theta}^0)$, and is equal to zero based on the definition of $\hat{\theta}^0$, but the last term, $c$, remains positive. It implies that $W$ can be further increased by increasing $\hat{\theta}$ from $\hat{\theta}^0$: although a more conservative threshold would deter otherwise productive reform, it further deters the detrimental reform that is undertaken by the low type by decreasing $\rho^*$. The reduced loss could more than compensate for the sacrificed gain from those otherwise efficient reforms with value in $(\theta^0, \hat{\theta}^*)$. Therefore, the social optimum must require a cautious attitude about potential reform, despite it inhibiting seemingly beneficial reform.

Reform can be permitted, i.e., $\hat{\theta}^* < \theta_2$, if and only if condition (7) is met. Because $\rho$ decreases with $\alpha$ (by Proposition 1), the left hand side of (7) strictly decreases with $\alpha$. Hence, this condition is more likely to be met with a larger $\alpha$, i.e., the presence of a higher proportion of high-talent politicians in the population. When the talent required for successful reform is
very scarce, the public would not expect sufficient gain from reform. The public then prefer no reform at all.

Similarly, the condition is more likely to be met with a larger $q$. In other words, reform is socially beneficial only when the success of reform is sufficiently likely.

These arguments further lead to more general conclusions on the impact of $\alpha$ and $q$ on the properties of $\hat{\theta}^* \in (\hat{\theta}_0, \theta_2)$.

**Proposition 5.** The socially optimal cutoff $\hat{\theta}^*$ decreases with $\alpha$ and $q$.

A greater $\alpha$ or $q$ always allows for less restriction on the politician’s activities.

**Example: Uniform Distribution**

Proposition 3 demonstrates that the equilibrium behaviour depends on the properties of the distribution of $\theta$. We now discuss its impact on the optimal threshold rule $\hat{\theta}^*$. To allow for a handy and informative analysis, consider an example in which the value of reform follows a uniform distribution

$$F(\theta) = \frac{\theta + \theta_1}{\theta_2 + \theta_1}$$

and the high-talent politician receives a perfect signal with $q = 1$.

An increase in $\theta_2$ implies that the probability mass of the distribution is shifted upward, high-valued reform proposals are more likely to occur, and more beneficial opportunities can be expected. The environment thus seems to favor more reform. Before we examine its impact on the socially optimal institutional rule, let us examine its welfare implications in an arbitrary equilibrium with a fixed $\hat{\theta}$. Figure 1 testifies to a non-monotonic relationship between social welfare and $\theta_2$ when $\hat{\theta}$ is given. The society may not be better off when more opportunities are available. The logic can be seen in Proposition 3: for a given cutoff $\hat{\theta}$, a stochastically dominant distribution of $\theta$ forces the low type to reform more, which increases the loss from his inefficient reform.

The ambiguous welfare implication compels us to further look into its implications on the socially optimal institution $\hat{\theta}^*$. The implications of a higher $\theta_2$ on the social optimum $\hat{\theta}^*$ are also ambiguous. On the one hand, low-valued reform proposals would emerge less often, and cause less damage, which encourages a more liberal rule to reap more benefits from reform. On the other hand, it could demand a more conservative rule in order to discipline the low type. Our analysis leads to the following proposition.

**Proposition 6.** The socially optimal cutoff point for reform, $\hat{\theta}^*$, strictly increases with $\theta_2$. 

19
Figure 1: An example that demonstrates the non-monotonic effect of $\theta_2$ on social welfare ($\theta_1 = 1.1$, $q = 1.0$, $\alpha = 0.2$, and $\bar{\theta} = 1.2$).

We find that when the probability mass of the uniform distribution is shifted upward, i.e., when more opportunities for reform can be expected, it unambiguously lifts the optimal cutoff $\hat{\theta}^*$. That is, a more optimistic environment requires additional caution and a more conservative socially optimal rule.

5 Discussion and Extensions

Our model exhibits a few distinct characteristics. Our paper is closely related to Majumdar and Mukand (2004), Suurmond, Swank, and Visser (2004), and Chen (2010). As discussed in Section 1, our paper differs from these papers in both focuses and settings.

In the models of Majumdar and Mukand (2004) and Suurmond, Swank, and Visser (2004), the performance of the risky project is depicted by a binary indicator (success or failure). The likelihood of its success is pre-determined while a more capable agent can discover the pre-determined “suitability” of the project more precisely. In contrast, we assume that (1) the performance of reform is a continuous measure, depending on the quality of both the project per se and the implementation by the politician; (2) the politician’s ability determines the quality of ex post implementation; and (3) the quality of the project is random.
Our model has a closer relation to that of Chen (2010). In both studies, the likelihood of success of the risky project is assumed to depend on the agent’s ability. Our setup and that of Chen (2010), however, differ from each other in several other aspects. First, in the setup of Chen (2010), the likelihood of success also depends on a random variable, whose realization is observable only to the agent, while the principal’s payoffs from success or failure of the risky project are prefixed. In contrast, in our model, the public’s payoffs from a successful or failed reform are dependent on a random variable, whereas the success and failure probabilities are affected by the politician’s ability alone. Second, in Chen’s (2010) model, even if the risky project fails, the agent’s reputation is still higher than that from choosing the safe project. In our model, the politician’s reputation from a failed reform is lower than that from choosing the status quo.

These diverse modeling approaches serve the differing research focuses of these studies. The papers thus complement each other. In our context, the setup enriches our analysis in two aspects. First, it enables an analysis of institution design. A more extensive trade-off is involved in determining the proper level of institutional conservatism. Second, a comparative static analysis may be performed on the probability distribution of the value of reform, which sheds further light on the equilibrium behaviour and the corresponding design of welfare-maximizing institution.

For the sake of expositional efficiency and mathematical tractability, our analysis has been limited to a stylized setting. It leaves open many possibilities for extensions and variations. For instance, the model may be extended to allow for a larger strategy space, or to allow the payoff of the politician to depend on the realized outcome of his policy choice. The analysis may also be extended to a dynamic setting. For example, the “pressure to prove oneself” result points toward the following conjecture: a politician who has failed in the past is more likely to take radical action in the future. Past failure lowers his rating among the public, which therefore makes it more lucrative for him to pursue accidental success in the future. Although these extensions would not yield predictions that fundamentally depart from our main results, they may yet spawn richer comparative statics that further add to our understanding of this issue. In the remainder of this paper, we discuss specifically two straightforward extensions of our model: (1) a setting where the politician also values his policy performance $y$; and (2) a setting where the success or failure of a reform may not be perfectly observed by the public.

\footnote{In her model, the agent is prevented from choosing the risky project only by monetary incentives.}
5.1 When the Politician Values Policy Performance

Assume that the politician not only cares about his reputation payoff, but also receives utility from the output of his policy choice. His objective function is written generically as

$$u(y, i) = \delta \Pr(t = H|y, i) + (1 - \delta)y,$$

with $\delta \in [0, 1]$. Let $\rho_t(\theta)$ denote the behavioral strategy of a type-$t$ politician, which indicates the probability he undertakes reform when a reform proposal of value $\theta$ is received. If the politician maintains the status quo, the output remains zero. His expected payoff is given by

$$u_0 = \alpha F(\hat{\theta}) + \alpha \int_\hat{\theta}^{\theta_2} [1 - \rho_H(\theta)] f(\theta) \, d\theta + (1 - \alpha) \int_\theta^{\theta_2} [1 - \rho_L(\theta)] f(\theta) \, d\theta.$$

The payoff $u_0$ does not differ from that in the basic model. By contrast, if the politician implements a reform of value $\theta$, his ends up with a payoff

$$u_s(\theta) = \delta \frac{\alpha q \rho_H(\theta) f(\theta)}{\alpha q \rho_H(\theta) f(\theta) + (1 - \alpha) \frac{1}{2} \rho_L(\theta) f(\theta)} + (1 - \delta)\theta$$

if the reform succeeds; and

$$u_f(\theta) = \delta \frac{\alpha (1 - q) \rho_H(\theta) f(\theta)}{\alpha (1 - q) \rho_H(\theta) f(\theta) + (1 - \alpha) \frac{1}{2} \rho_L(\theta) f(\theta)} + (1 - \delta)(\theta - 4),$$

if it fails.

A smaller $\delta$ implies that the politician is subject to weaker reputation concerns. The model boils down to the first best benchmark when $\delta$ reduces to zero; while it approximates our original model when $\delta$ approaches one. We characterize the equilibrium of the game with a given $\hat{\theta}$ in the following proposition. For expositional efficiency, we consider only the case of $\hat{\theta} \geq 4(1 - q).

Remark 1. 1. When $\delta$ is sufficiently small, i.e., when the condition $\frac{\delta}{1 - \delta} \frac{1 - \alpha}{\alpha F(\hat{\theta}) + (1 - \alpha)} \leq 2 - \theta_2$ is met, the high-type politician reforms with probability one for all $\theta \geq \hat{\theta}$, while the low-type politician does not reform in the equilibrium of the game.

2. When $\delta$ is sufficiently large, i.e., when the condition $\frac{\delta}{1 - \delta} \frac{1 - \alpha}{\alpha F(\hat{\theta}) + (1 - \alpha)} > 2 - \theta_2$ is met, the high-type politician reforms with probability one for all $\theta \geq \hat{\theta}$ in the equilibrium of the game. There exists a unique cutoff $\tilde{\theta}_L \in [\hat{\theta}, \theta_2)$, such that the low type reforms with a positive probability $\rho_L^*(\theta)$ for all $\theta > \tilde{\theta}_L$, with $\rho_L^*(\theta|\hat{\theta})$ strictly increasing with $\theta$.

The proof is similar to that for Proposition 1. We omit for brevity but it is available from the author upon request. The equilibrium of the game ultimately depends on the size
of $\delta$. When the politician’s utility also depends on the actual output $y$, the politician bears additional loss from his unsuccessful reform, which may discourage low-type politician from undertaking reform. When $\delta$ is sufficiently small, the strong incentives from output $y$ may even lead to the possibility of full separation between types. However, whenever nontrivial reputation concerns are present, i.e. $\frac{\delta}{1-\delta} \frac{1-\alpha}{\alpha F(\hat{\theta})+1-\alpha} > 2 - \theta_2$, equilibrium behaviour resembles those in the basic model. The low type mimics his high-type counterpart, despite that he plays a strictly monotone equilibrium with $\rho^*_L(\theta|\hat{\theta})$ strictly increasing with $\theta$.

The trade-off on social welfare remains in the extended setting. We present the following remark.

**Remark 2.** Under nontrivial reputation concerns, i.e. $\frac{\delta}{1-\delta} \frac{1-\alpha}{\alpha F(\hat{\theta})+1-\alpha} > 2 - \theta_2$, the cutoff $\bar{\theta}_L$ strictly increases with $\hat{\theta}$. For all $\theta \in [\hat{\theta}, \theta_2]$, $\rho^*_L(\theta|\hat{\theta})$ strictly decreases with $\hat{\theta}$.

**Proof.** See Appendix. ☐

Remark 2 demonstrates that the low-type politician would reform less when a more stringent standard is in place. With nontrivial reputation concerns, institutional conservatism can be still in demand as it deters the inefficient reform conducted by the low type.

## 6 Concluding Remarks

In this paper, we study a politician’s incentive to implement reform when his true ability is privately known but he is concerned about the public’s perception of his abilities. The politician thus chooses his policy and action to maximize his reputation payoff. We find that a high-talent politician always attempts to reform as much as possible, which compels his low-talent counterpart to mimic with a positive probability. Socially inefficient reform therefore results. Further, we explore the socially optimal level of empowerment, and find that the social optimum can be achieved only if the prevailing institutional rule implements proper conservatism and deters some otherwise efficient reform.

### Appendix: Proofs

**Proof of Proposition 1**

**Divinity Criterion**

We first formally translate the notion of the Divinity Criterion into our context. Fix an equilibrium with a cutoff $\bar{\theta} > \hat{\theta}$. Suppose that an unexpected reform with a value $\theta \in [\hat{\theta}, \bar{\theta})$ takes place. The public infers from its outcome the value of $\theta$. The public forms a set
of beliefs $\phi_\theta \equiv \{\tilde{\rho}_H(\theta), \tilde{\rho}_L(\theta)\}$, where $\tilde{\rho}_t(\theta)$ specifies the probability of a type-$t$ politician to undertake this reform. Given this conjecture, a type-$t$ politician, when deviating, has a payoff

$$\mu_t(\theta; \phi_\theta) = q_t \times \frac{\alpha \tilde{\rho}_H(\theta)q}{\alpha \tilde{\rho}_H(\theta)q + \frac{1}{2}(1 - \alpha)\tilde{\rho}_L(\theta)} + (1 - q_t) \times \frac{\alpha \tilde{\rho}_H(\theta)(1 - q)}{\alpha \tilde{\rho}_H(\theta)(1 - q) + \frac{1}{2}(1 - \alpha)\tilde{\rho}_L(\theta)}.$$

Let $\mu^*_t$ denote the payoff of a type-$t$ politician in the equilibrium. Further define $\Phi^t_\theta \equiv \{\phi_\theta | \mu_t(\theta; \phi_\theta) > \mu^*_t\}$. We then have the following.

**Definition 1.** Under Divinity Criterion, the out-of-equilibrium belief $\phi_\theta$ satisfies:

$$\tilde{\rho}_t(\theta) \geq \tilde{\rho}_{t'}(\theta) \text{ if } \Phi^t_\theta \subset \Phi^{t'}_\theta, \text{ with } t \in \{H, L\} \text{ and } t \neq t'.$$

We claim the following results.

**Claim 1.** Suppose that in a hypothetical equilibrium in which there exists $\theta \in [\tilde{\theta}, \theta_2]$, with $\rho_t(\theta) = 0$, $\forall t \in \{L, H\}$. Then $\Phi^L_\theta \subset \Phi^H_\theta$.

**Proof.** Consider a hypothetical deviation of a reform with value $\theta$. Define $\tilde{\alpha} \equiv \frac{\alpha \tilde{\rho}_H(\theta)}{\alpha \tilde{\rho}_H(\theta) + (1 - \alpha)\tilde{\rho}_L(\theta)}$.

The high type, if deviates, has an _ex ante_ expected payoff

$$\mu_H(\theta; \tilde{\alpha}) = q \times \frac{\tilde{\alpha}q}{\tilde{\alpha}q + \frac{1}{2}(1 - \tilde{\alpha})} + (1 - q) \times \frac{\tilde{\alpha}(1 - q)}{\tilde{\alpha}(1 - q) + \frac{1}{2}(1 - \tilde{\alpha})} = q \times \frac{1}{1 + \frac{1}{2\tilde{\alpha}q}(1 - \tilde{\alpha})} + (1 - q) \times \frac{1}{1 + \frac{1}{2\tilde{\alpha}(1 - q)}(1 - \tilde{\alpha})}.$$

She has an incentive to deviate if and only if $\pi_H(\theta) - \mu^0 \geq 0$. The low type, by contrast, has an _ex ante_ expected payoff

$$\mu_L(\theta; \tilde{\alpha}) = \frac{1}{2} \times \frac{1}{1 + \frac{1}{2\tilde{\alpha}q}(1 - \tilde{\alpha})} + \frac{1}{2} \times \frac{1}{1 + \frac{1}{2\tilde{\alpha}(1 - q)}(1 - \tilde{\alpha})}.$$

She has an incentive to deviate if and only if $\pi_L(\theta) - \mu^0 \geq 0$. Because $\frac{1}{1 + \frac{1}{2\tilde{\alpha}q}(1 - \tilde{\alpha})} > \frac{1}{1 + \frac{1}{2\tilde{\alpha}(1 - q)}(1 - \tilde{\alpha})}$, we see that $\mu_H(\theta) - \mu^0 > 0$ whenever $\mu_L(\theta) - \mu^0 \geq 0$. It implies that the high type is always more likely to deviate by undertaking an expected reform than the low type. 

Claim 1 demonstrates that the high type always benefits more from reform.
Equilibrium Characterization

Claim 2. There exists no equilibrium with $\theta \in [\hat{\theta}, \theta_2]$, and $\rho_t(\theta) = 0$, $\forall t \in \{L, H\}$.

Proof. The out-of-equilibrium belief must require $\tilde{\alpha} \geq \alpha$ to reflect the result of Claim 1. We now prove $\mu_H(\theta) > \alpha$. To see this, observe that

$$
\mu_H(\theta; \tilde{\alpha}) = q \times \frac{\tilde{\alpha} q}{\tilde{\alpha} q + \frac{1}{2} (1 - \tilde{\alpha})} + (1 - q) \times \frac{\tilde{\alpha} (1 - q)}{\tilde{\alpha} (1 - q) + \frac{1}{2} (1 - \tilde{\alpha})}
$$

$$
> \frac{\tilde{\alpha} q}{\frac{1}{2} (1 - \tilde{\alpha})} \left[ \frac{\tilde{\alpha} q + \frac{1}{2} (1 - \tilde{\alpha})}{\tilde{\alpha} q + \frac{1}{2} (1 - \tilde{\alpha})} \right]

+ \frac{\tilde{\alpha} (1 - q)}{\frac{1}{2} (1 - \tilde{\alpha})} \left[ \frac{\tilde{\alpha} (1 - q) + \frac{1}{2} (1 - \tilde{\alpha})}{\tilde{\alpha} (1 - q) + \frac{1}{2} (1 - \tilde{\alpha})} \right]

= \tilde{\alpha} > \alpha,
$$

where we have used the fact that $q > 1/2$. Given such a belief, the high type must deviate when $\theta$ is realized, because his expected payoff $\mu_H(\theta) > \alpha > \mu^0$. The original equilibrium cannot be sustained by a belief system that satisfies Divinity. ■

Claim 3. For any $\theta \in [\hat{\theta}, \theta_2]$, in an equilibrium,

1. if the low type reforms with positive probability, the high type must reform with probability one;
2. if the high type does not reform, the low type would not reform with positive probability;
3. $\rho_H(\theta) > 0$, $\forall \theta \in [\hat{\theta}, \theta_2]$.

Proof. When the politician does not undertake a reform, his expected payoff $\mu^0$ is independent of his type. Whenever he reform, the expected payoff for a high type is $q \mu^s + (1 - q) \mu^f$, which is higher than that for the low type, $\frac{1}{2} \mu^s + \frac{1}{2} \mu^f$. Hence, if we have $\frac{1}{2} \mu^s + \frac{1}{2} \mu^f \geq \mu^0$, then $q \mu^s + (1 - q) \mu^f > \mu^0$. Further, if we have the high type choose not to reform, i.e. $q \mu^s + (1 - q) \mu^f \leq \mu^0$, then $\frac{1}{2} \mu^s + \frac{1}{2} \mu^f < \mu^0$.

Suppose that there exists $\theta \in [\hat{\theta}, \theta_2]$ with $\rho_H(\theta) > 0$. By Claim 2 $\rho_L(\theta) > 0$. Contradiction. ■

Claim 4. In the equilibrium, the politician must play montone strategy, such that $\rho_t(\theta)$ must be nondecreasing with $\theta$ for $\theta \in [\hat{\theta}, \theta_2]$, $\forall t \in \{L, H\}$.

Proof. Suppose that there exist $\theta, \theta' \in [\hat{\theta}, \theta_2]$, with $\theta > \theta'$ and $\rho_t(\theta) < \rho_t(\theta')$. Recall the definition of $q_t$. We must have

$$
q_t \mu^s(\theta) + (1 - q_t) \mu^f(\theta) \leq q_t \mu^s(\theta') + (1 - q_t) \mu^f(\theta'),
$$

25
which is written as
\[
q_t \frac{\alpha q q_H(\theta)}{\alpha q q_H(\theta) + (1 - \alpha)\frac{1}{2} q_L(\theta)} + (1 - q_t) \frac{\alpha(1 - q) q_H(\theta)}{\alpha(1 - q) q_H(\theta) + (1 - \alpha)\frac{1}{2} q_L(\theta)} \\
\leq q_t \frac{\alpha q q_H(\theta')}{\alpha q q_H(\theta') + (1 - \alpha)\frac{1}{2} q_L(\theta')} + (1 - q_t) \frac{\alpha(1 - q) q_H(\theta')}{\alpha(1 - q) q_H(\theta') + (1 - \alpha)\frac{1}{2} q_L(\theta')}
\]
\[
(11)
\]

By Claim 2, \( q_H(\cdot) \) cannot be zero. The condition is further rewritten as
\[
q_t \left[ \frac{\alpha q}{\alpha q + (1 - \alpha)\frac{1}{2} q_L(\theta)} - \frac{\alpha q}{\alpha q + (1 - \alpha)\frac{1}{2} q_L(\theta')} \right] \\
\leq (1 - q_t) \left[ \frac{\alpha(1 - q)}{\alpha(1 - q) + (1 - \alpha)\frac{1}{2} q_L(\theta)} - \frac{\alpha(1 - q)}{\alpha(1 - q) + (1 - \alpha)\frac{1}{2} q_L(\theta')} \right],
\]
which requires \( \frac{q_L(\theta)}{q_H(\theta)} \geq \frac{q_L(\theta')}{q_H(\theta')} \).

Suppose \( q_L(\theta), q_L(\theta') > 0 \), then by Claim 3 \( q_H(\theta) = q_H(\theta') = 1 \). Then we have \( q_L(\theta) \geq q_L(\theta') \). Contradiction.

Suppose \( q_L(\theta) = q_L(\theta') = 0 \). In that case, undertaking reform gives a payoff of one, which is not an equilibrium, as the low type must deviate. Contradiction. \( \square \)

**Claim 5.** \( q_H(\theta) = 1 \) and \( q_L(\theta) > 0 \), \( \forall \theta \in [\hat{\theta}, \theta_2] \).

**Proof.** We claim that whenever the high type chooses reform with a positive probability, the low type must do so as well. We have shown that whenever both types choose reform with positive probability, the high type’s probability of reform is one and therefore at least as high as the low type’s. Therefore, the overall probability for the low type to choose the status quo, \( P_{0L} \), is weakly higher than that for the high type, \( P_{0H} \). Thus, if the low type chooses the status quo, his reputation is \( \mu = \frac{\alpha q \rho_H(\theta)}{\alpha q \rho_H(\theta) + (1 - \alpha)\frac{1}{2} q_L(\theta)} \leq \alpha \).

However, if he deviates and undertakes reform, he is believed to be a high type with probability one if \( q < 1 \). If \( q = 1 \), his payoff depends on the public’s off-equilibrium belief when reform fails. However, he succeeds with probability \( \frac{1}{2} \), and the resulting expected payoff still exceeds \( \alpha \). Therefore, it cannot be that the low type always chooses the status quo when the high type chooses reform.

By Claim 3, \( q_H(\theta) > 0 \), \( \forall \theta \in [\hat{\theta}, \theta_2] \). Then \( q_L(\theta) > 0 \), \( \forall \theta \in [\hat{\theta}, \theta_2] \). Again, by Claim 3, \( q_H(\theta) = 1 \), \( \forall \theta \in [\hat{\theta}, \theta_2] \). \( \square \)

### 6.0.1 Proof of Proposition 1

We now determine the low-type politician’s probability of reform for a proposal with value \( \theta \), which we denote by \( q(\theta) \) to economize on notation. By (3), if the politician maintains the
status quo, his payoff is

\[
\mu^0 = \frac{\alpha F(\hat{\theta})}{\alpha (1 - \alpha) F(\hat{\theta}) + (1 - \alpha) f(\hat{\theta}) [1 - \rho(\hat{\theta})]} \cdot \frac{\hat{\theta}}{F(\hat{\theta})}.
\]

(12)

Note that it does not depend on \( \theta \). On the other hand, if the low-type politician undertakes the reform, his payoff is given by

\[
\mu_L(\theta) = \frac{1}{2} \cdot \frac{q(\alpha f(\theta) + \frac{1}{2} (1 - \alpha) \rho(\theta) f(\theta)) + (1 - q)(1 - \alpha) \rho(\theta) f(\theta)}{\alpha + \frac{1}{2} (1 - \alpha) \rho(\theta) + \frac{1}{2}}.
\]

(13)

If the low-type plays a completely mixed strategy, \( \rho(\theta) \in (0, 1) \), we need to equate (12) and (13), which implies that \( \rho(\theta) \) must be a constant \( \rho \) regardless of the value \( \theta \). Consequently, in equilibrium,

\[
\alpha \frac{\alpha}{\alpha + (1 - \alpha) \frac{F(\hat{\theta}) + (1 - \alpha) \rho(\theta) f(\hat{\theta})}{F(\hat{\theta})}} = \frac{1}{2} \cdot \frac{\alpha}{\alpha + \frac{1}{2} (1 - \alpha) \rho(\theta) + \frac{1}{2}}.
\]

(14)

which we may rewrite as

\[
\frac{1}{1 + \lambda(\alpha) A} = \frac{1}{1 + \lambda(\alpha) B} + \frac{1}{1 + \lambda(\alpha) C},
\]

(15)

where

\[
\lambda(\alpha) = \frac{1 - \alpha}{\alpha}, \quad A = 1 + (1 - \rho) \kappa(\hat{\theta}), \quad \kappa(\hat{\theta}) = \frac{1 - F(\hat{\theta})}{F(\hat{\theta})}, \quad B = \frac{1}{2} \rho, \quad C = \frac{1}{2} \rho.
\]

This is the same equation as (3). The expression \( \lambda(\alpha) \) is the likelihood ratio of the low type versus the high type, \( \kappa(\hat{\theta}) \) is the likelihood ratio of reform having good prospects versus bad prospects, and \( A, B, \) and \( C \) are respectively the likelihood ratios of the low type not reforming, having a successful reform, and having a failed reform versus the high type obtaining each outcome. Consider the equilibrium condition (15). Note that its \( LHS \) is \( \mu^0 \) and its \( RHS \) is \( \mu_L \). When \( \rho = 0 \), \( \mu^0 \leq \alpha \), while \( \mu_L = 1 \) as \( B = C = 0 \). Therefore, \( \mu^0 < \mu_L \). By contrast, when \( \rho = 1 \), \( \mu^0 = \alpha \) as \( A = 1 \), and \( \mu_L < \alpha \), which can be seen from the fact that when \( \rho = 1 \)

\[
\alpha \mu_H + (1 - \alpha) \mu_L = \alpha,
\]

while \( \mu_L < \mu_H \). Therefore, \( \mu^0 > \mu_L \).

Both the \( RHS \) and \( LHS \) of (15) are continuous in \( \rho \). Furthermore, it is straightforward to show that the \( LHS \) strictly increases with \( \rho \), while the \( RHS \) strictly decreases with \( \rho \). Hence, we conclude that there must exist a unique \( \rho^* \in (0, 1) \) that solves (15).
Proof of Proposition 2

We first establish the following.

Claim 6. In the equilibrium, $\rho^*$ strictly decreases with $\hat{\theta}$.

Recall the equilibrium condition (16). When $\hat{\theta}$ increases, $\kappa(\hat{\theta}) \equiv \frac{1-F(\hat{\theta})}{F(\hat{\theta})}$ must decrease, which causes $g(\rho^*, \alpha, q, \hat{\theta})$ to decrease. Further, as we have shown in the proof for previous results, $g(\rho^*, \alpha, q, \hat{\theta})$ strictly decreases with $\rho^*$. By the implicit function theorem, we establish that when $\hat{\theta}$ increases, $\rho^*$ must decrease.

We then use the result to establish the main claim of Proposition 3.

Recall that the equilibrium is defined by the equation

$$
\frac{\alpha}{1 + (1-\alpha)(1-e^{\lambda}\frac{1}{1-F(\hat{\theta})})} = \frac{\alpha}{\alpha + (1-\alpha)\frac{1}{1-q}} \mu^* + \frac{\alpha}{\alpha + (1-\alpha)\frac{1}{1-q}} \mu^f.
$$

The politician in office receives a payoff $\mu^0$ when he maintains the status quo. He receives a payoff $\mu^s$ when he successfully implements a reform and $\mu^f$ when he fails. In any equilibrium with a given $\hat{\theta}$, the type-$t$ politician receives a payoff

$$
\mu_t = \begin{cases} 
q_t\mu^s + (1-q_t)\mu^f, & \text{for } \theta \geq \hat{\theta}; \\
\mu^0, & \text{for } \theta < \hat{\theta}.
\end{cases}
$$

Hence, in this equilibrium, the expected payoff of a type-$t$ politician is given by

$$
E(\mu_t) = \mu^0 F(\hat{\theta}) + [q_t\mu^s + (1-q_t)\mu^f][1 - F(\hat{\theta})].
$$

First, we claim that when $\hat{\theta}$ increases, $E(\mu_H)$ and $E(\mu_L)$ change in opposite directions. Therefore, the first part of the proposition implies the second part. This claim is an implication of the fact $\alpha E(\mu_H) + (1-\alpha)E(\mu_L) = \alpha$, or

$$
E(\mu_H) = 1 - \lambda(\alpha)E(\mu_L).
$$

Now, we prove the first part of the proposition. For a low-type politician, $E(\mu_L) = \mu^0$ because $\mu^0 = \frac{1}{2}\mu^s + \frac{1}{2}\mu^f$. Hence, we need only verify $\frac{d\mu^0}{d\theta} > 0$. Define

$$
H(\rho^*, \hat{\theta}) = \frac{\alpha}{1 + (1-\alpha)(1-e^{\lambda}\frac{1}{1-F(\hat{\theta})})} - \frac{1}{2}\left[\frac{\alpha}{\alpha + (1-\alpha)\frac{1}{1-q}} \mu^* + \frac{\alpha}{\alpha + (1-\alpha)\frac{1}{1-q}} \mu^f\right].
$$

We have

$$
\frac{d\mu^0}{d\theta} = \frac{\partial\mu^0}{\partial\theta} + \frac{\partial\mu^0}{\partial\rho^*} \cdot \frac{\partial\rho^*}{\partial\theta} = \frac{\partial\mu^0}{\partial\theta} + \frac{\partial\mu^0}{\partial\rho^*} \cdot \left[-\frac{\partial H(\rho^*, \hat{\theta})}{\partial \rho^*}\right].
$$

Because $\frac{\partial H(\rho^*, \hat{\theta})}{\partial \rho^*} = \frac{\partial\mu^0}{\partial\theta}$, we then have $\frac{d\mu^0}{d\theta} = \frac{\partial\mu^0}{\partial\theta}[1 + \frac{\partial H(\hat{\theta}, \rho^*)}{\partial \rho^*}]$. We must have $1 - \frac{\partial H(\rho^*, \hat{\theta})}{\partial \rho^*} > 0$ because $\frac{\partial H(\rho^*, \hat{\theta})}{\partial \rho^*} = \frac{\partial\mu^0}{\partial\theta} - \frac{1}{2}\left(\frac{\partial \mu^s}{\partial\rho^*} + \frac{\partial \mu^f}{\partial\rho^*}\right)$, while $\frac{\partial\mu^0}{\partial\theta} > 0$, $\frac{\partial \mu^s}{\partial\rho^*}, \frac{\partial \mu^f}{\partial\rho^*} < 0$. 

28
Proof of Proposition 3

Part 1 Consider the equilibrium condition (15). We have shown above that the left hand side of (15) is increasing in $\rho^*$ and the right hand side decreasing in $\rho^*$. Note that $A$, $B$, and $C$ do not contain $\alpha$ in their expressions. Thus, we may write

$$
\frac{\partial(LHS - RHS) \text{ of (15)}}{\partial \alpha} = -\frac{1}{\alpha^2} \left[ -\frac{A}{(1 + \lambda(\alpha)A)^2} + \frac{1}{2} \cdot \frac{B}{(1 + \lambda(\alpha)B)^2} + \frac{1}{2} \cdot \frac{C}{(1 + \lambda(\alpha)C)^2} \right].
$$

We want to evaluate the above derivative at the value of $\rho$ that satisfies (15). Observe that $0 < B < C$ as $q \geq 3/4 > 1/2$, we may conclude then $B < A < C$ based on (15). From (15), we obtain

$$
\frac{A}{1 + \lambda(\alpha)A} = \frac{1}{2} \cdot \frac{B}{1 + \lambda(\alpha)B} + \frac{1}{2} \cdot \frac{C}{1 + \lambda(\alpha)C}.
$$

Therefore,

$$
\frac{1}{2} \cdot \frac{B}{(1 + \lambda(\alpha)B)^2} + \frac{1}{2} \cdot \frac{C}{(1 + \lambda(\alpha)C)^2} = \frac{A}{1 + \lambda(\alpha)A} \left[ \frac{B}{1 + \lambda(\alpha)B} + \frac{C}{1 + \lambda(\alpha)C} \cdot \frac{1}{1 + \lambda(\alpha)B} + \frac{C}{1 + \lambda(\alpha)C} \cdot \frac{1}{1 + \lambda(\alpha)C} \right].
$$

The expression in the brackets is a convex combination of $\frac{1}{1 + \lambda(\alpha)B}$ and $\frac{1}{1 + \lambda(\alpha)C}$. Since $0 < B < C$, the former is larger, but the coefficient on the former is smaller than $\frac{1}{2}$. Using (15), we have

$$
\frac{1}{2} \cdot \frac{B}{(1 + \lambda(\alpha)B)^2} + \frac{1}{2} \cdot \frac{C}{(1 + \lambda(\alpha)C)^2} < \frac{A}{(1 + \lambda(\alpha)A)^2}.
$$

Hence, at the value of $\rho$ that satisfies (15),

$$
\frac{\partial(LHS - RHS) \text{ of (15)}}{\partial \alpha} > 0.
$$

Thus, by the implicit function theorem, the probability of reform by the low type, $\rho^*$, is decreasing in $\alpha$, the probability of high type.

Next, we verify the comparative statics of $\bar{\rho}$. Because $\frac{\partial \rho^*}{\partial \alpha} < 0$, we only need to show $\left| (1 - \alpha) \frac{\partial \rho^*}{\partial \alpha} \right| + \rho^* > 1$. We have

$$
\left| (1 - \alpha) \frac{\partial \rho^*}{\partial \alpha} \right| + \rho^* = \frac{\rho^* \left[ 1 - \frac{1}{4q(1-q)} \left( \frac{\partial \rho^*}{\partial \alpha} \right)^2 \right]}{\left[ \kappa(\hat{\theta}) + 1 + \frac{4q(1-q)}{4q(1-q) + \lambda(\alpha)\rho^*} \right]} + \rho^*.
$$
Rearranging the equilibrium condition leads to

\[(1 - \rho^*) \kappa(\hat{\theta}) = \frac{\rho^*(\lambda(\alpha)\rho^* + 1)}{4q(1 - q) + \lambda(\alpha)\rho^*} - 1\]

\[= \frac{\rho^*(\lambda(\alpha)\rho^* + 1) - 4q(1 - q) - \lambda(\alpha)\rho^*}{4q(1 - q) + \lambda(\alpha)\rho^*}\]

\[= \frac{\lambda(\alpha)\rho^{*2} + \rho^* - \lambda(\alpha)\rho^* - 4q(1 - q)}{4q(1 - q) + \lambda(\alpha)\rho^*},\]

which yields

\[\kappa(\hat{\theta}) = \frac{\lambda(\alpha)\rho^{*2} + \rho^* - \lambda(\alpha)\rho^* - 4q(1 - q)}{4q(1 - q) + \lambda(\alpha)\rho^*}(1 - \rho^*),\]

and therefore

\[\kappa(\hat{\theta}) + 1 = \frac{\lambda(\alpha)\rho^{*2} + \rho^* - \lambda(\alpha)\rho^* - 4q(1 - q) + [4q(1 - q) + \lambda(\alpha)\rho^*](1 - \rho^*)}{4q(1 - q) + \lambda(\alpha)\rho^*}(1 - \rho^*)\]

\[= \frac{\rho^*[1 - 4q(1 - q)]}{4q(1 - q) + \lambda(\alpha)\rho^*}(1 - \rho^*).\]

Hence,

\[\left[ \kappa(\hat{\theta}) + 1 + \frac{4q(1 - q)[1 - 4q(1 - q)]}{4q(1 - q) + \lambda(\alpha)\rho^*} \right]\]

\[= \frac{\rho[1 - 4q(1 - q)]}{4q(1 - q) + \lambda(\alpha)\rho^*}(1 - \rho^*) + \frac{4q(1 - q)[1 - 4q(1 - q)]}{4q(1 - q) + \lambda(\alpha)\rho^*}^2\]

\[= \frac{1 - 4q(1 - q)}{4q(1 - q) + \lambda(\alpha)\rho^*}^2(1 - \rho^*)[4q(1 - q) + \lambda(\alpha)\rho^{*2}].\]

We then obtain

\[\left| (1 - \alpha) \frac{\partial \rho^*}{\partial \alpha} \right| + \rho^*\]

\[= \frac{(1 - \alpha)}{\alpha^2} \cdot \frac{\rho^{*2}[1 - 4q(1 - q)]}{4q(1 - q) + \lambda(\alpha)\rho^*}^2 + \rho^*\]

\[= \frac{(1 - \alpha)}{\alpha^2} \cdot \frac{(1 - \rho^*)\rho^{*2}}{4q(1 - q) + \lambda(\alpha)\rho^*} + \rho^*.\]

For our purpose, we only need to show \(\frac{(1 - \alpha)}{\alpha^2} \cdot \frac{\rho^{*2}[1 - 4q(1 - q)]}{4q(1 - q) + \lambda(\alpha)\rho^*}^2 > 1.\) Rewrite it as \(\frac{(1 - \alpha)}{\alpha^2} \cdot \frac{\lambda(\alpha)\rho^{*2}}{4q(1 - q) + \lambda(\alpha)\rho^*} = \frac{1}{\alpha^2} \cdot \frac{\lambda(\alpha)\rho^{*2}}{4q(1 - q) + \lambda(\alpha)\rho^*} + \frac{1}{\lambda(\alpha)\rho^{*2} + 1}.\)

Hence, it suffices to show \(\frac{1}{\lambda(\alpha)\rho^{*2} + 1} > \alpha.\)

We claim \(\frac{1}{\lambda(\alpha)\rho^{*2} + 1} > \frac{1}{2} > \alpha,\) i.e., \(4q(1 - q) > \lambda(\alpha)\rho^{*2}.\) To show that, recall the equilibrium condition \(1 + (1 - \rho^*)m = \frac{\rho^*(\lambda(\alpha)\rho^* + 1)}{4q(1 - q) + \lambda(\alpha)\rho^*},\) which implies \(\frac{\rho^*(\lambda(\alpha)\rho^* + 1)}{4q(1 - q) + \lambda(\alpha)\rho^*} > 1 \Rightarrow \rho^*(\lambda(\alpha)\rho^* + 1) > 4q(1 - q) + \lambda(\alpha)\rho^* \Leftrightarrow \lambda(\alpha)\rho^{*2} + \rho^* > 4q(1 - q) + \lambda(\alpha)\rho^*.\) Because \(\lambda(\alpha) > 1, \lambda(\alpha)\rho^{*2} > 4q(1 - q)\) must hold.
Part 2 The equilibrium condition can be rewritten as

\[ g(\rho^*, \alpha, q, \hat{\theta}) \equiv [1 + (1 - \rho^*)\kappa(\hat{\theta})] - \frac{\rho^*[\lambda(\alpha)\rho^* + 1]}{4q(1 - q) + \lambda(\alpha)\rho} = 0. \] (16)

Since \( q \geq \frac{3}{4} \), \( G(\rho^*, \alpha, q) \) is decreasing with \( q \). Further,

\[ \frac{\partial g(\rho^*, \alpha, q, \hat{\theta})}{\partial \rho^*} = -\left[ \kappa(\hat{\theta}) + 1 + \frac{4q(1 - q)[1 - [4q(1 - q)]^2]}{[4q(1 - q) + \lambda\rho^*]^2} \right] < 0. \]

Recall that \( \kappa(\hat{\theta}) = \left[ 1 - F(\hat{\theta}) \right]/F(\hat{\theta}) \). We then obtain

\[ \frac{d\rho^*}{dq} = -\frac{\partial g(\rho^*, q)}{\partial q} \frac{\partial q}{\partial \rho^*} < 0. \]

That \( \bar{\rho} \) is decreasing in \( q \) is an immediate consequence by its definition in (4).

Part 3 Consider the equilibrium condition (15). Since \( F \) first order stochastically dominates \( G \), we have \( F(\hat{\theta}) < G(\hat{\theta}) \). This implies that for any given \( \rho \), LHS of (15) for \( F \) is lower than that for \( G \), since \( \kappa(\hat{\theta}) \) is larger for \( F \) than for \( G \).

As we have shown above, LHS of (15) strictly increases with \( \rho \), while RHS strictly decreases. Thus, only if \( \rho > \rho' \) can make (15) hold for both distributions. From this, the definition of \( \bar{\rho} \) in (4), and the assumption that \( F \) first order stochastically dominates \( G \), we can immediately see that \( \bar{\rho} > \bar{\rho}' \).

Proof of Lemma 1

Consider the value of

\[ E(y|\hat{\theta}, \hat{\theta}) = \alpha[\hat{\theta} - 4(1 - q)] + (1 - \alpha)\rho^*(\hat{\theta} - 2). \]

When \( \hat{\theta} = 4(1 - q) \), it must be negative. When \( \hat{\theta} \) approaches \( \theta_2 \), we have its value approach \( \alpha[\theta_2 - 4(1 - q)] + (1 - \alpha)\rho(\theta_2 - 2) \), which is positive if and only if \( \frac{(1 - \alpha)\rho}{\alpha} < \frac{\theta_2 - 4(1 - q)}{2 - \theta_2} \). Further recall that \( E(y|\theta, \hat{\theta}) \) strictly increases with both \( \theta \) and \( \hat{\theta} \). There must exist a unique \( \hat{\theta} \) that solves the equation.

Lemma 2 and Its Proof

Because \( f(\hat{\theta}) > 0 \) for all \( \hat{\theta} \in [-\theta_1, \theta_2] \), the sign of (9) is the same as that of \( \frac{dW}{d\hat{\theta}}/f(\hat{\theta}) \). For our purpose, it suffices to explore \( \frac{dW}{d\hat{\theta}}/f(\hat{\theta}) \). We then establish the following lemma.

Lemma 2. The expression \( \frac{dW}{d\hat{\theta}}/f(\hat{\theta}) \) strictly decreases with \( \hat{\theta} \).
Recall that the equilibrium condition (16) with a given \( \hat{\theta} \) can be written as
\[
g(\rho^*, \kappa) = [1 + (1 - \rho^*) \kappa(\hat{\theta})] - \frac{\rho^*(\lambda \rho^* + 1)}{4q(1 - q) + \lambda \rho^*} = 0,
\]
where \( \kappa(\hat{\theta}) = [1 - F(\hat{\theta})]/F(\hat{\theta}) \), as defined in (15). Hence, we have \( \frac{\partial g(\rho^*, \kappa(\hat{\theta}))}{\partial \rho^*} = \lambda(1 - \rho^*) \).
Because \( \frac{\partial g(\rho^*, \kappa(\hat{\theta}))}{\partial \rho^*} = -[\kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-4q(1-q)^2]}{[4q(1-q) + \lambda \rho^*]^2}] < 0 \), we must have
\[
\frac{d\rho^*}{d\kappa} = \frac{1 - \rho^*}{\kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-4q(1-q)^2]}{[4q(1-q) + \lambda \rho^*]^2}},
\]
and therefore
\[
\frac{d\rho^*}{d\theta} / f(\hat{\theta}) = -\frac{1 - \rho^*}{[\kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-4q(1-q)^2]}{[4q(1-q) + \lambda \rho^*]^2}][F(\hat{\theta})]^2}.
\]
We now claim \( \frac{d\rho^*}{d\theta} / f(\hat{\theta}) \) strictly decreases with \( \hat{\theta} \). We have
\[
d\left[\frac{d\rho^*}{d\theta} / f(\hat{\theta})\right] = \left[ \frac{\frac{\partial}{\partial \theta} \left[\kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-4q(1-q)^2]}{[4q(1-q) + \lambda \rho^*]^2}\right][F(\hat{\theta})]^2}{\left[\kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-4q(1-q)^2]}{[4q(1-q) + \lambda \rho^*]^2}\right][F(\hat{\theta})]^2} \right] \cdot \frac{1}{1 - \rho^*}
\]
Note that \( \frac{d\rho^*}{d\theta} [\kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-4q(1-q)^2]}{[4q(1-q) + \lambda \rho^*]^2}][F(\hat{\theta})]^2 = (1 - \rho^*) f(\hat{\theta}) \). We then only need to prove
\[
d\left[\frac{\partial}{\partial \theta} \left[\kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-4q(1-q)^2]}{[4q(1-q) + \lambda \rho^*]^2}\right][F(\hat{\theta})]^2 \right] > f(\hat{\theta}).
\]
Rewrite \( \kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-4q(1-q)^2]}{[4q(1-q) + \lambda \rho^*]^2} \) as \( \frac{4q(1-q)[1-4q(1-q)^2]}{[4q(1-q) + \lambda \rho^*]^2} \) and \( F(\hat{\theta}) \) strictly increases. Hence, \( \frac{d\left[\kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-4q(1-q)^2]}{[4q(1-q) + \lambda \rho^*]^2}\right][F(\hat{\theta})]^2}{d\theta} > 0 \). Furthermore, \( \frac{dF(\hat{\theta})}{d\theta} = f(\hat{\theta}) \). We then establish our claim.

**Proof of Proposition 4**

If \( \frac{(1-\alpha)\rho}{\alpha} \geq \frac{\theta_2-4(1-q)}{2-\theta_2} \), then \( \hat{\theta}^0 \) does not exist. Any reform with a value \( \hat{\theta} < \theta_2 \) must lead to negative expected outcome. Hence, no reform is \textit{ex ante} beneficial, which implies \( \hat{\theta}^* = \theta_2 \). If \( \frac{(1-\alpha)\rho}{\alpha} < \frac{\theta_2-4(1-q)}{2-\theta} \), then \( \hat{\theta}^0 \) exists. \( \frac{dW}{d\theta} / f(\hat{\theta}) \left|_{\hat{\theta} = \hat{\theta}^0} > 0 \right. \) but \( \frac{dW}{d\theta} / f(\hat{\theta}) \left|_{\hat{\theta} = \theta_2} < 0 \right. \) (because \( \frac{(1-\alpha)\rho}{\alpha} < \frac{\theta_2-4(1-q)}{2-\theta} \)), then there must exist a unique \( \hat{\theta}^* \in (\hat{\theta}^0, \theta_2) \) that solves \( \frac{dW}{d\theta} / f(\hat{\theta}) = 0 \).

**Proof of Proposition 5**

Suppose that an interior optimum with \( \hat{\theta}^* \in (0, \theta_2) \) exists. Define \( k \equiv \left[-\frac{d\rho^*}{d\theta} / f(\hat{\theta})\right] \). Then the optimal condition is
\[
v \equiv \alpha[\hat{\theta} - 4(1-q)] + (1-\alpha)\rho^*(\hat{\theta} - 2) - (1-\alpha)k \int_{\hat{\theta}}^{\theta_2} (2-\theta) f(\theta) d\theta = 0. \quad (17)
\]
Apparently, \( \frac{dv}{d\theta} = -\frac{dw}{d(\theta^2)} > 0 \). We now claim \( \frac{dv}{da} > 0 \). Taking first order derivative of \( v \) yields

\[
\frac{dv}{d\alpha} = \left[ \hat{\theta} - 4(1 - q) \right] - \rho^*(\hat{\theta} - 2) + (1 - \alpha) \frac{d\rho^*}{d\alpha} (\hat{\theta} - 2) + k \int_{\theta_1}^{\theta_2} (2 - \theta) f(\theta) d\theta - (1 - \alpha) \frac{dk}{d\alpha} \int_{\theta_1}^{\theta_2} (2 - \theta) f(\theta) d\theta.
\]

It suffices to show \( k \) strictly decreases with \( \alpha \) and \( q \). Recall by the proofs of previous results:

\[
- \frac{d\rho^*}{d\alpha} = \frac{\rho^*[1 - 4q(1 - q)]}{4q(1 - q) + \lambda \rho^*} \cdot \left| \frac{d\lambda(\alpha)}{d\alpha} \right|.
\]

Note \( -\frac{d\rho^*}{d\alpha} = \frac{-d\rho^*}{d\theta} \). Hence, we now evaluate \( -\frac{d\rho^*}{d\theta} \) with respect to \( \hat{\theta} \). We first rearrange it as

\[
- \frac{d\rho^*}{d\alpha} = \frac{\rho^*[1 - 4q(1 - q)]}{4q(1 - q) + \lambda \rho^*} \cdot \left| \frac{d\lambda(\alpha)}{d\alpha} \right|.
\]

By the proof of part 1 of Proposition 3, \( \frac{(1 - \rho^*)}{[\kappa(\hat{\theta}) + 1 + \frac{4q(1 - q)}{4q(1 - q) + \lambda \rho^*}]^2} \) decreases with \( \hat{\theta} \). We claim \( \frac{\rho^2}{1 - \rho^*} \cdot \frac{1}{4q(1 - q) + \lambda \rho^*} \) also decreases with \( \hat{\theta} \). Evaluate it with respect to \( \hat{\theta} \) yields

\[
\frac{\rho^*(2 - \rho^*)}{(1 - \rho^*)^2} \cdot \frac{1}{4q(1 - q) + \lambda \rho^*}^2 - 2\lambda \frac{d\rho^*}{d\theta} \frac{\rho^*}{1 - \rho^*} \cdot \frac{1}{4q(1 - q) + \lambda \rho^*}^2.
\]

Because \( \frac{d\rho^*}{d\theta} < 0 \), we need to show \( (2 - \rho^*)[4q(1 - q) + \lambda \rho^*] - 2\lambda \rho^*(1 - \rho^*) > 0 \), which is obvious because \( (2 - \rho^*)[4q(1 - q) + \lambda \rho^*] - 2\lambda \rho^*(1 - \rho^*) = (2 - \rho^*)[4q(1 - q) + \lambda \rho^*] - \lambda \rho^*(2 - 2\rho^*) \), and \( 2 - \rho^* > 2 - 2\rho^* \).

We further claim \( \hat{\theta}^* \) decreases with \( q \). To show that, we have to prove \( \frac{dv}{dq} > 0 \). We have

\[
\frac{dv}{dq} = 4\alpha + (1 - \alpha) \frac{d\rho^*}{dq} (\hat{\theta} - 2) - (1 - \alpha) \frac{dk}{dq} \int_{\theta_1}^{\theta_2} (2 - \theta) f(\theta) d\theta.
\]

It would suffice to show \( \frac{dk}{dq} < 0 \). We use the same technique as above. We have

\[
- \frac{d\rho^*}{dq} = \frac{\rho^*[1 - 4q(1 - q)]}{4q(1 - q) + \lambda \rho^*} \cdot \left| \frac{d\lambda(\alpha)}{d\alpha} \right|.
\]
We then claim \(-\frac{\partial^2 \nu^*}{\partial \phi \partial \theta} < 0\). Rewrite \(-\frac{d \rho^*}{dq}\) as
\[
\frac{d \rho^*}{dq} = \frac{1 - \rho^*}{\kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-[4q(1-q)]^2]}{[4q(1-q) + \lambda \rho^*]^2}} \cdot \frac{1}{1 - \rho^*} \cdot \frac{4(2q - 1)\rho^*(\lambda \rho^* + 1)}{[4q(1-q) + \lambda \rho^*]^2}.
\]
Because \(\frac{1 - \rho^*}{\kappa(\hat{\theta}) + 1 + \frac{4q(1-q)[1-[4q(1-q)]^2]}{[4q(1-q) + \lambda \rho^*]^2}}\) and \(\frac{1}{1 - \rho^*}\) decreases with \(\hat{\theta}\), we only need to show \(\frac{\rho^*(\lambda \rho^* + 1)}{[4q(1-q) + \lambda \rho^*]^2}\) decreases with \(\hat{\theta}\). Taking first order derivative of it with respect to \(\hat{\theta}\) yields
\[
\frac{d \rho^*(\lambda \rho^* + 1)}{d \theta} = \left[ \frac{(2\lambda \rho^* + 1)\frac{d \rho^*}{d \theta} [4q(1-q) + \lambda \rho^*]^2(1 - \rho^*)}{4q(1-q) + \lambda \rho^*} - 2\rho^*(\lambda \rho^* + 1)(1 - \rho^*)[4q(1-q) + \lambda \rho^*] \lambda \frac{d \rho^*}{d \theta} \right] \cdot \left(1 - \rho^*\right)^2 [4q(1-q) + \lambda \rho^*]^4
\]
By Lemma 2, \(\frac{d \rho^*}{d \theta} < 0\). Hence, it remains to verify that the item in bracket is positive. This is obvious because
\[
\left[ (2\lambda \rho^* + 1)[4q(1-q) + \lambda \rho^*](1 - \rho^*) - 2\rho^*(\lambda \rho^* + 1)(1 - \rho^*) \right] \cdot \left(1 - \rho^*\right)^2 [4q(1-q) + \lambda \rho^*]^4 > \lambda \rho^* [(2\lambda \rho^* + 1)(1 - \rho^*) - 2(\lambda \rho^* + 1)(1 - \rho^*) + \rho^*(\lambda \rho^* + 1)]
\]
= \(\lambda \rho^* [-\rho^* + \rho^*(\lambda \rho^* + 1)] > 0\).

**Proof of Proposition 6**

We examine how a higher upper support \(\theta_2\) could affect \(dW/d\hat{\theta}\) for any given \(\hat{\theta}\). When the high-type politician is perfectly informed, a closed form for \(\rho^*\) is obtained as
\[
\rho^* = 1 - \frac{\alpha}{1 - \alpha} F(\hat{\theta})..
\]
The first-order derivative of the welfare function is derived as follows
\[
\frac{dW}{d\hat{\theta}} = \frac{1}{\theta_2 + \hat{\theta}} \left\{ -\alpha \hat{\theta} - (1 - \alpha) \rho^*(\hat{\theta} - 2) \right. + (1 - \alpha) \frac{d \rho^*}{d \hat{\theta}} \int_{\hat{\theta}}^{\theta_2} (\theta - 2) f(\theta) d\theta \}
\]
\[
= \frac{1}{\theta_2 + \hat{\theta}} \left\{ -\alpha \hat{\theta} + (1 - \alpha) \left(1 - \frac{\alpha(\hat{\theta} + \theta_1)}{(1 - \alpha)(\theta_2 + \theta_1)}\right)(2 - \hat{\theta}) + \frac{\alpha(\theta_2 - \hat{\theta})}{2}[4 - (\theta_2 + \hat{\theta})] \right. \}
\]
The optimal cutoff \(\hat{\theta}^*\) is determined by the equation
\[
v = -\alpha \hat{\theta} + (1 - \alpha) \left(1 - \frac{\alpha(\hat{\theta} + \theta_1)}{(1 - \alpha)(\theta_2 + \theta_1)}\right)(2 - \hat{\theta}) + \frac{\alpha(\theta_2 - \hat{\theta})}{2}[4 - (\theta_2 + \hat{\theta})] = 0.
\]
By Lemma 4, \( v \) strictly decreases with \( \hat{\theta} \). We only need to show \( \frac{\partial v}{\partial \theta} > 0 \). Apparently, \( (1 - \alpha)[1 - \frac{\alpha(\hat{\theta} + \hat{\theta})}{(1 - \alpha)\alpha(\theta + \hat{\theta})}](2 - \hat{\theta}) \) increases with \( \theta_2 \). We claim \( \frac{\alpha(\hat{\theta} + \hat{\theta})}{(1 - \alpha)(\theta_2 + \hat{\theta})} \) increases with it as well. Taking first order derivative of \( (\theta_2 - \hat{\theta})[4 - (\theta_2 + \hat{\theta})] \) yields \( 4 - (\theta_2 + \hat{\theta}) - (\theta_2 - \hat{\theta}) = 4 - 2\theta_2 > 0 \), which completes the proof.

### 6.1 Proof of Remark 2

Define \( \hat{\theta}_L = \min(\theta | \rho_L(\theta, \hat{\theta}) > 0) \). Whenever \( \rho_L(\theta, \hat{\theta}) > 0 \), the equilibrium is determined by the system of equations:

\[
\delta \left[ \frac{1}{2} \frac{\alpha q + (1 - \alpha) \frac{1}{2} \rho_L(\theta, \hat{\theta})}{\alpha (1 - q)} + \frac{\alpha}{\alpha (1 - q) + (1 - \alpha) \frac{1}{2} \rho_L(\theta, \hat{\theta})} \right] + (1 - \delta)(\theta - 2) = \frac{\alpha F(\hat{\theta})}{\alpha F(\hat{\theta}) + (1 - \alpha) - (1 - \alpha) \int_{\theta_L}^{\theta_2} \rho_L(\theta_0, \hat{\theta}) f(\theta) d\theta}, \forall \theta \in [\theta_L, \theta_2],
\]

\[
\delta \left[ \frac{1}{2} \frac{\alpha q + (1 - \alpha) \frac{1}{2} \rho_L(\theta, \hat{\theta})}{\alpha (1 - q)} + \frac{\alpha}{\alpha (1 - q) + (1 - \alpha) \frac{1}{2} \rho_L(\theta, \hat{\theta})} \right] + (1 - \delta)(\theta - 2) = \delta \left[ \frac{1}{2} \frac{\alpha q + (1 - \alpha) \frac{1}{2} \rho_L(\theta, \hat{\theta})}{\alpha (1 - q)} + \frac{\alpha}{\alpha (1 - q) + (1 - \alpha) \frac{1}{2} \rho_L(\theta, \hat{\theta})} \right] + (1 - \delta)(\theta' - 2), \forall \theta, \theta' \in [\theta_L, \theta_2].
\]

We first prove that \( \rho_L^*(\theta, \hat{\theta}) \) strictly decreases with \( \hat{\theta} \) by contradiction. By the equilibrium condition, for arbitrary \( \theta, \theta' \in [\theta_L, \theta_2] \), if \( \rho_L(\theta, \hat{\theta}) \) in(de)creases, then \( \rho_L(\theta', \hat{\theta}) \) must in(de)crease as well.

Suppose that \( \hat{\theta} \) drops but \( \rho_L(\theta, \hat{\theta}) \) decreases. To have the low type reform less, the payoff for no reform, i.e.

\[
\frac{\alpha F(\hat{\theta})}{\alpha F(\hat{\theta}) + (1 - \alpha) - (1 - \alpha) \int_{\theta_L}^{\theta_2} \rho_L(\theta_0, \hat{\theta}) f(\theta) d\theta},
\]

must strictly increase. Because \( F(\hat{\theta}) \) decreases, \( (1 - \alpha) - (1 - \alpha) \int_{\theta_L}^{\theta_2} \rho_L(\theta, \hat{\theta}) f(\theta) d\theta \) must decrease, which requires \( \int_{\theta_L}^{\theta_2} \rho_L(\theta, \hat{\theta}) f(\theta) d\theta \) to increase. Let \( \hat{\theta} \) drop to \( \theta \). We consider the following possibilities.

**Case 1: \( \theta_L \) also increases.** Then \( \int_{\theta_L}^{\theta_2} \rho_L(\theta, \hat{\theta}) f(\theta) d\theta \) must decrease. Contradiction.

**Case 2: \( \theta_L \) decreases.** Let \( \theta' \) drop to \( \theta_L \). There are altogether four possibilities.

**Case 2.1 \( \theta_L > \theta \) and \( \theta' \) increases.** Consider \( \theta \in (\theta_L, \theta') \). Before the drop, the low type does not want to undertake the reform for such \( \theta \). Hence, we have \( \delta + (1 - \delta)(\theta - 2) < \frac{\alpha F(\hat{\theta}) + (1 - \alpha) - (1 - \alpha) \int_{\theta_L}^{\theta_2} \rho_L(\theta_0, \hat{\theta}) f(\theta) d\theta}{\alpha F(\hat{\theta}) + (1 - \alpha) - (1 - \alpha) \int_{\theta_L}^{\theta_2} \rho_L(\theta, \hat{\theta}) f(\theta) d\theta} \). After the drop, the low type reforms with a positive probability in the equilibrium. He receives \( \delta \left[ \frac{1}{2} \frac{\alpha q + (1 - \alpha) \frac{1}{2} \rho_L(\theta', \hat{\theta})}{\alpha (1 - q)} + \frac{\alpha(1 - q)}{\alpha (1 - q) + (1 - \alpha) \frac{1}{2} \rho_L(\theta', \hat{\theta})} \right] + (1 - \delta)(\theta - 2) \), which is less than \( \delta + (1 - \delta)(\theta - 2) \). However, we also have \( \delta \left[ \frac{1}{2} \frac{\alpha q + (1 - \alpha) \frac{1}{2} \rho_L(\theta_0, \hat{\theta})}{\alpha (1 - q)} + \frac{\alpha(1 - q)}{\alpha (1 - q) + (1 - \alpha) \frac{1}{2} \rho_L(\theta_0, \hat{\theta})} \right] + (1 - \delta)(\theta - 2) \).
\[
\frac{\alpha F'(\hat{\theta})}{\int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta} + \frac{\alpha F'_{\theta}(\hat{\theta})}{\int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta}, \]
which is required to be greater than
\[
\frac{\alpha F'(\hat{\theta})}{\int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta}.
\]

Contradiction.

**Case 2.2** $\bar{\theta}_L = \hat{\theta}$ and $\bar{\theta}_L' = \hat{\theta}'$. The payoff from no reform is written as
\[
\alpha + (1 - \alpha) \frac{1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}') f(\theta) d\theta}{F(\hat{\theta})} = \alpha + (1 - \alpha) \frac{1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}') f(\theta) d\theta}{F(\hat{\theta})}.
\]

Further, when $\hat{\theta}$ drops, the denominator becomes
\[
\alpha + (1 - \alpha) \frac{1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}') f(\theta) d\theta}{F(\hat{\theta})}.
\]
Before $\hat{\theta}$ drops, $\rho_L(\cdot | \hat{\theta}) < 1$, $\int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta < F(\hat{\theta}) - F(\hat{\theta}')$, which is defined as $\Delta$. Then
\[
\frac{1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}') f(\theta) d\theta}{F(\hat{\theta}) - \Delta} = \frac{1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}') f(\theta) d\theta}{F(\hat{\theta}) - \Delta} > 1 - \Delta
\]

The inequality holds if and only if
\[
1 - \Delta < \frac{1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}') f(\theta) d\theta}{1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta}.
\]

Because $\rho_L(\theta | \hat{\theta}') < \rho_L(\theta | \hat{\theta})$, $1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta > 1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}') f(\theta) d\theta$. Hence, it suffices to show
\[
\frac{\int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta}{1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta} < \frac{\Delta}{F(\hat{\theta})}.
\]
This is obvious, because $1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta > F(\hat{\theta})$ and $2 \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}') f(\theta) d\theta < \Delta$. Hence, the payoff from no reform $\alpha + (1 - \alpha) \frac{1 - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}) f(\theta) d\theta - \int_{\hat{\theta}}^{\theta} \rho_L(\theta | \hat{\theta}') f(\theta) d\theta}{F(\hat{\theta})}$ must decrease. Hence, the payoff from no reform must decrease instead of increasing. Contradiction.

**Case 2.3** $\bar{\theta}_L = \hat{\theta}$ and $\bar{\theta}_L > \hat{\theta}'$. The proof is similar to that for Case 2.2.

**Case 2.4** $\bar{\theta}_L > \hat{\theta}$ and $\bar{\theta}_L = \hat{\theta}'$. The proof is similar to that for Case 2.1.

We conclude that $\rho^*_L(\theta | \hat{\theta})$ strictly decreases with $\hat{\theta}$ must decrease with $\hat{\theta}$ for all $\theta \in [\bar{\theta}_L, \theta_2]$. Further, we establish that $\bar{\theta}_L$ must decrease with $\hat{\theta}$. Again, we prove the claim by contradiction. Assume that $\hat{\theta}$ decreases to $\hat{\theta}'$. Suppose the contrary that $\bar{\theta}_L$ increases to $\hat{\theta}_L'$. Because the low type reforms more for all $\theta \in [\bar{\theta}_L, \theta_2]$, the low type must have a lower expected payoff for all $\theta$, regardless of undertaking the reform or maintaining status quo. Consider an arbitrary $\theta = \bar{\theta}_L - \varepsilon > \bar{\theta}_L$, where $\varepsilon$ is an infinitely small positive number. In the new equilibrium, the low type prefers to receive the payoff of no reform for $\theta$, which must be higher than he would deviate. However, the payoff if he deviates must be strictly higher than his payoff in the previous equilibrium, as the public would believe he is the high type with certainty. This contradicts with the fact that the equilibrium payoff in the new equilibrium is lower than that of the previous equilibrium.

36
References


Skocpol, Theda, “Bringing the State Back In: Strategies of Analysis in Current Research,” in Peter B. Evans, Dietrich Rueschemeyer, and Theda Skocpol, eds., Bringing...
