A model of public opinion management

Andrea Patacconi[*]   Nick Vikander[†]

October 2014

Abstract

Policymakers often motivate their decisions using information collected by government agencies. While more information can help hold the government to account, it may also give policymakers an incentive to meddle with the work of bureaucrats. This paper develops a model of biased information gathering to examine how different disclosure rules and the degree of independence of government agencies affect citizen welfare. Disclosure rules and agency independence interact in subtle ways. We find that secrecy is never optimal and yet insulating the agency from political pressure, so that its information is always unbiased, may also not be socially optimal. A biased information gathering process can benefit the government by helping it to shape public opinion. But it can also benefit the public, by curbing the government's tendency to implement its ex ante favored policy, thus mitigating the agency conflict between policymakers and the public.

Keywords: Transparency, Accountability, Independence, Manipulation of Information.

JEL Classification: D73, H11, H56.

---

[*]Norwich Business School, University of East Anglia. Email: a.patacconi@uea.ac.uk

[†]University of Copenhagen. Email: nick.vikander@econ.ku.dk

Human experience teaches us that those who expect public dissemination of their remarks may well temper candor with a concern for appearances and for their own interest to the detriment of the decision-making process. (U.S. Supreme Court, United States v. Nixon)

We also recognize that there is a real dilemma between giving the public an authoritative account of the intelligence picture and protecting the objectivity of the JIC [Joint Intelligence Committee] from the pressures imposed by providing information for public debate. (Butler Report, p. 114)

1 Introduction

Transparency is an essential feature of a democratic and accountable state and yet, despite substantial progress in recent years, exceptions to the principle of open government remain commonplace (Prat, 2006). In the United States, for instance, the President has the right to withhold information from Congress and the courts, typically on the grounds that he needs candid and confidential advice from his staff. Freedom of Information laws also frequently allow policymakers to withhold information, most notably to protect internal decision making, personal privacy and national security (Banisar, 2004; Roberts, 2006).

This paper examines one important rationale for lack of transparency in government: the concern that public dissemination of information might compromise the quality of government decision making. We develop a model where the government receives information from an agency about a particular policy, and then decides whether or not the policy should be implemented. For instance, the government might receive an intelligence report about the opportunity to go to war, or an environmental impact assessment about the opportunity to build a new nuclear power plant. As is standard in political agency models, the preferences of the government and the public are not perfectly aligned. The government is more favorable than the public towards implementation but also wants public support for its decision. Thus, while policymakers may be more willing to wage war than voters, they are nevertheless responsive to public opinion.

Our key assumption is that the agency may be politicized and hence its report to the government may be biased. If the agency is independent, then it provides an unbiased report about the consequences of implementing the policy, and hence about the appropriate course of action. However, if the agency is *not* independent, then this report may be biased in favor of the government's ex ante preferred decision; that is, the report may be biased in favor of implementation. With a nonindependent agency, we assume that the government can choose

the optimal degree of bias so as to maximize its own welfare. For instance, the government may staff the agency with individuals who are prone to stating a case for war, seek the advice of biased experts, or encourage biased information gathering and evaluation. The drawback is that all parties with access to the report (including the government) then receive lower quality information which can result in poor decision making.

We use this framework to address two questions, both from the perspective of the public. First, should the contents of the report be publicly disclosed? And second, should the agency be made independent of the government? Both issues are of great practical importance. It is often claimed that secrecy is instrumental in protecting the integrity of the decision-making process and indeed one of the most common exemptions to the principle of open government concerns pre-decision information (Banisar, 2004). Granting independence to government agencies is also becoming increasingly common. The Federal Trade Commission in the U.S. and the Bank of England, for instance, have a status that ensures their independence from political pressure by limiting the removal of their heads to certain specific causes. The British commission in charge of investigating recent episodes of intelligence failure also recommended to strengthen the independence of the Joint Intelligence Committee, although it fell well short of recommending full independence from the executive (Butler Report, 2004, pp. 143-144).

In line with conventional wisdom, we find that disclosure ('transparency') makes the government more accountable and hence more responsive to public desires, relative to nondisclosure ('secrecy'). However, disclosure also induces policymakers to distort the process of information gathering and evaluation. In contrast, when no information can be disclosed, the government has no incentive to manipulate information. Secrecy is therefore effective at protecting the integrity of the decision-making process.

We also consider a constitutional stage in which both the disclosure rule and the agency's degree of independence can be specified. The most surprising results emerge regarding what rule and degree of independence maximize the public's welfare. We show that from the public's perspective, secrecy is *never* optimal, but it can be optimal for the government agency *not* to be independent. Secrecy is always dominated by transparency because its chief advantage – unbiased information – can be more efficiently obtained by insulating the agency from political pressure. And yet the public may sometimes prefer that the agency be politicized so that its report is potentially biased. The government may opt for a biased agency that tends to submit favorable reports, because these reports help shape public opinion. However,

for any given decision rule, biased information increases the probability that the government will make the wrong decision, which hurts both itself and the public. The government wants to avoid making the wrong decision and so taylors its optimal decision rule to the agency's level of bias. We show that a pro-implementation bias in information has a *moderating* effect; for given evidence, it makes the government more reluctant to implement the policy. This moderating effect benefits the public, which views implementation less favorably than the government. Thus, manipulation of information can help mitigate the agency conflict between the government and the public.

From a theoretical perspective, this result can be seen as an application of the theory of the second-best. According to this theory, introducing a new inefficiency – manipulation of information – in an environment where another inefficiency is already present – the agency conflict between the government and the public – can sometimes increase social welfare.

Previous work has examined how politicians can be held accountable when voters are not perfectly informed. Canes-Wrone et al. (2001) and Maskin and Tirole (2004) study models where policymakers have private information and reelection concerns create incentives for pandering. However, because these models do not allow policymakers to credibly communicate their private information to voters, they cannot distinguish between transparency and secrecy. Subsequent research has relaxed the assumption that voters are imperfectly informed by examining the role of media. Besley and Prat (2006) develop a model where incumbents (good and bad) can manipulate media reports by offering some form of compensation to the media owners. Their analysis focuses on how features of the media industry affect the quality of the media reports and political turnover. Ashworth and Shotts (2010) find that a government can have lower incentives to pander when the media has a tendency to act as a "yes-man" (thus herding on the incumbent's choice) because negative media reports then become strongly indicative of an incorrect policy choice. Warren (2012) focuses on the motivations of news-media providers. He shows that a moderate degree of pro-incumbent bias helps improve government accountability because, if in equilibrium even bad incumbents act in the public interest, then journalists that exert investigative effort will tend to uncover positive information about the incumbents. As a consequence, a pro-incumbent bias gives journalists an extra incentive to work hard.

There are several important differences between the present paper, Ashworth and Shotts (2010), and Warren (2012). In our model, information is collected by an agency and communicated to the government before a decision is taken; moreover, the quality of the information gathering-process can be affected by the government. In Ashworth and Shotts (2010) and Warren (2012), the role of news-media providers is simply to verify ex post what information the government had when the decision was taken. These papers provide valuable insights into the motivations and behavior of the media.[1] However, they do not address the issue of how to protect the quality of government information. They also do not examine questions of institutional design such as the choice between transparency and secrecy or between independent and nonindependent bureaucrats.

One important similarity between our work, Ashworth and Shotts (2010), and Warren (2012), is that in all three settings, some degree of pro-incumbent bias can improve social welfare. However, the mechanism driving this result is completely different in each setting. In Ashworth and Shotts (2010), the rational tendency of news-media providers to herd on the incumbent's choice makes negative reports more informative. In Warren (2012), a pro-incumbent bias strengthen news-media providers' incentives. In our model, biased information has a moderating influence on government policy and can more closely align policymakers' and citizens' preferences. These papers are thus best seen as complements.

The present paper is also related to the literature on transparency in principal-agent relationships (e.g., Holmstrom, 1979; Cremer, 1995; Stasavage 2004). Prat (2005), in particular, develops a model of career concerns for experts where the principal can observe the agent's action and/or its consequences. He shows that transparency on action can induce the agent to disregard useful private information and act in a conformist manner. As a consequence, the principal can be better off by committing not to observe the action. Transparency on consequences, by contrast, always benefits the principal. Fox (2007) develops a related model where the concern for policymakers is not to prove that they are competent, but to show that they are unbiased. The present paper differs from these articles because it focuses neither on transparency on action nor on consequences. We measure transparency by the extent to which *pre-decision information* is shared between the agent and the principal. Our focus is

---

[1]For theoretical analyses of bureaucratic behavior, see also Prendergast (1993, 2007), Gailmard and Patty (2007), Alesina and Tabellini (2007, 2008), Patty (2009), Shotts and Wiseman (2010) and Ujhelyi (2014). Our analysis leaves the motivations of agency bureaucrats in the background, allowing us to focus on how disclosure rules and public opinion can shape policy. Interactions between politicians, bureaucrats and voters are explored in Fox and Jordan (2011).

not whether transparency induces conformism on the part of the agent, but whether an agent will distort his own information (and possibly the principal's) to influence how the principal perceives his action.[2]

The remainder of the paper is organized as follows. The next section introduces the model. Sections 3 and 4 study different disclosure rules (transparency and secrecy), under the assumption that the government agency is nonindependent. Section 5 considers the case of an independent agency and compares different institutional arrangements from the public's point of view. Extensions are discussed in Section 6, while Section 7 concludes. Proofs are gathered in two technical appendices.

2   Model

We consider a model of government decision making where (i) the government is responsive to public opinion and (ii) the agency that provides the government with information is potentially biased. The model has four stages. At stage 1, if the agency is nonindependent, then the government chooses the agency's level of bias, $q \in [0, 1]$. One can interpret q as the type of bureaucrats who work at the agency. In contrast, if the agency is independent, its bias is equal to zero ($q \equiv 0$). At stage 2, the agency produces a report for the government. This report may or may not be publicly revealed, depending on the disclosure rule, as discussed below. At stage 3, the government and the public sequentially take actions. First, the government must choose whether to implement a new policy ($p = a$) or stick with the status quo ($p = n$). If the government decides to implement the new policy, then the public can protest ($v = d$) or accept ($v = nd$) implementation. The cost to the public of protesting is $c \geq 0$; the benefit is that implementation fails with probability p, in which case the policy reverts to the status quo. Thus $p \in (0, 1)$ is a measure of the power of public opinion at constraining government action. When p is close to 1, the public can almost always block the implementation of the new policy. If the government selects the status quo, we assume that the public does not protest, and that the status quo remains.[3] At stage 4, payoffs are

---

[2]Levy (2007) and Swank et al. (2008) develop models closely related to Prat's to study the effect of transparency on committee decision making. They show that secrecy can be conducive to better decision making because, if individual votes cannot be observed, then voters have less of an incentive to distort their actions in order to signal their types.

[3]This assumption is without loss of generality, because the government is biased in favor of implementation relative to the public (see below). Even if the public could protest against the status quo, to pressure the government to implement the new policy, it would never do so in equilibrium in our setting.

realized.

Preferences. The payoffs of the government and the public depend on the state of the world, $S \in \{A, N\}$. The public would like the policy to match the true state, $a = A$ or $n = N$, in which case its payoff is zero. The public incurs a loss of C_a if the policy is implemented and the true state is N, and a loss of C_n if the policy is not implemented and the true state is A. Without loss of generality, we assume that $C_a + C_n = 1$ and that A and N are a priori equally likely.[4]

Let σ_P denote the public's posterior belief that the true state is A and assume that the government implements the new policy. The public protests if its utility from protesting, $(1-p)[-C_a(1 - \sigma_P)] + p[-C_n\sigma_P] - c$, is higher than its utility from accepting implementation, $-C_a(1 - \sigma_P)$. This condition can be rewritten as

$$C_a - \sigma_P > \frac{c}{p}. \tag{1}$$

Thus, the public protests when protesting is effective (p large), when the cost of protesting, c, is small, when state A is believed to be unlikely (σ_P low) and when the cost of mistaken implementation, C_a, is relatively high. We assume that in a pure strategy equilibrium, if the public is indifferent, then it will accept implementation.[5]

The preferences of the government are not entirely congruent with those of public. Like the public, the government also incurs a loss of C_a from implementing the policy when the state is N, and a loss of C_n from not implementing the policy when the state is A. However, in addition to this concern for public welfare (a 'legacy' concern), the government also enjoys a private benefit $B \geq 0$ when the new policy is implemented.[6] Finally, the government suffers an explicit loss $L > 0$ whenever the public protests against its decision. The parameter L captures in a stylized fashion a number of costs associated with a loss of popularity or legitimacy, such as economic disruption caused by protests, vilification by the

---

[4]Assuming that A and N are equally likely is convenient, as it simplifies expressions for the posterior beliefs of the public and the government. Relaxing this assumption would be notationally burdensome but would not qualitatively affect the analysis.

[5]When the public is uncertain about a politician's type, fully rational Bayesian information processing can induce the public to vote for an incumbent who is known to have chosen a suboptimal policy (from the public's point of view). This is because the suboptimal policy may convey information about the politician's type, and voters want to select the best candidate going forward (see, e.g, Daley and Snowberg, 2009; Fox and Shotts, 2009; Ashworth, 2012). This issue does not arise in the present model, where there is no uncertainty about government skill or policy preferences.

[6]We assume that the public knows the value B. This is a reasonable assumption for situations where the government is well established and its policy preferences are widely known.

press, or greater likelihood that the government may step down in the future. Thus, L, like p, is a measure of how powerful public opinion is at disciplining the government.

Let σ_{Gov} denote the government's belief that $S = A$. Suppose that the public protests when the government selects implementation. Then the government payoff when choosing implementation is $(1 - p)[-C_a(1 - \sigma_{Gov}) + B] + p[-C_n\sigma_{Gov}] - L$, while it is $-C_n\sigma_{Gov}$ when it chooses the status quo. The government selects implementation over the status quo if

$$-C_a + \sigma_{Gov} + B \geq \frac{L}{1 - p}. \tag{2}$$

The government is more likely to select implementation when the public is unable to effectively discipline the government (L and p low), when the private benefits from implementation B are high, and when the cost to the public of mistaken implementation (the legacy concern) is relatively small (σ_{Gov} high and C_a low). We assume that in a pure strategy equilibrium, if the government is indifferent, then it will select implementation. In the analysis that follows, we define

$$E \equiv \frac{L}{1 - p}. \tag{3}$$

This parameter captures the strength of public pressure at constraining the government.[7]

Information Structure. Before making a policy decision, the government receives a report from the agency. This report is composed of two signals, $s_i \in \{\alpha, \varnothing\}$, $i = 1, 2$. A α signal provides evidence in support of implementation, while a $\varnothing$ signal provides evidence in support of the status quo. If the agency is independent, then these signals are genuine, s_i^G. Genuine signals are informative, conditionally independent and satisfy $\Pr(s_i^G = \alpha|A) = \Pr(s_i^G = \varnothing|N) = \theta$, where $\theta \in (\frac{1}{2}, 1)$ measures the signal precision.[8]

If the agency is nonindependent, then the signal-generating process may be distorted. Let $\mathbf{s}^q = \{s_1^q, s_2^q\}$ be the report produced by a nonindependent agency with bias $q \in [0, 1]$.

---

[7]To illustrate $p > 0$ and $L > 0$, take the 2012 Quebec student protests over provincial government plans to increase university tuition fees. After months of sustained protest, the government was voted out of office in September 2012, and the policy was reversed. The crisis was estimated to have cost the government $26 million for extra policing, as well as costing $33 million to the publicly-funded network of community colleges, and $20 million to Université du Québec à Montréal alone. For more details, see Chapter 3 of the exhaustive government report, "Rapport, Commission spéciale d'examen des événements du printemps 2012", Quebec Ministry for Public Security, March 2014, available in French at www.securitepublique.gouv.qc.ca. See also "UQAM claims student protests cost it $20 million" by Kevin Dougherty and Michelle Lalond, Montreal Gazette, October 30, 2012, and "Liberal government dismisses report on student unrest as a political attack" by the Canadian Press, May 15, 2014.

[8]We use two signals to allow for situations where the evidence is mixed. We use binary signals (instead of a single signal with multiple signal realizations) because this allows for a simple parametrization of the process of information manipulation, as explained below.

We capture the idea of *asymmetric vetting* by assuming that with probability q, a genuine $\varnothing$ signal is transformed into a fake α signal. That is, the nonindependent agency garbles the signal-generating process so that $\Pr(s_i^q = \alpha \mid s_i^G = \varnothing) = q$, which is independent across signals. The probability that a genuine α signal is transformed into a fake $\varnothing$ signal is zero, $\Pr(s_i^q = \varnothing \mid s_i^G = \alpha) = 0$. Thus q measures the agency's bias in favor of the government's ex ante preference for implementation. A non-independent agency with zero bias will behave just like an independent agency, and produce a report consisting of genuine signals, $\mathbf{s}^0 = \mathbf{s}^G = \left\{ s_1^G, s_2^G \right\}$.

If the agency is nonindependent, we allow the government to choose q to maximize its own payoff. This assumption is plausible if the government can appoint key agency personnel or can punish or reward them. We also posit that the government only observes the biased signals $\mathbf{s}^q$, rather than the genuine ones. This captures the fundamental drawback of manipulations: information is lost which may have been useful for decision making.[9]

Before proceeding, we introduce some additional notation. Let $(\cdot, \cdot)^q$ be a shorthand for $\mathbf{s}^q = (\cdot, \cdot)$. Any party that observes $\mathbf{s}^q$ will update its beliefs about the true state, where we define $\sigma_+^q \equiv \Pr(A|(\alpha, \alpha)^q)$, $\sigma^q \equiv \Pr(A|(\alpha, \varnothing)^q) = \Pr(A|(\varnothing, \alpha)^q)$, $\sigma_- \equiv \sigma_-^q \equiv \Pr(A|(\varnothing, \varnothing)^q) = \sigma_-^G$. These beliefs correspond to the three possible cases that can arise: (i) the report supports implementation ($\mathbf{s}^q = (\alpha, \alpha)$), (ii) the report is mixed ($\mathbf{s}^q = (\alpha, \varnothing)$ or $(\varnothing, \alpha)$) or (iii) the report supports the status quo ($\mathbf{s}^q = (\varnothing, \varnothing)$).[10] We sometimes refer to α signals as positive signals, and to $\varnothing$ signals as negative signals. It is easy to verify that rational agents discount the α signals more than the $\varnothing$ signals because the α signals can be forged: $\sigma^q \leq \sigma^G = \frac{1}{2}$ and $\sigma_+^q \leq \sigma_+^G$. This effect becomes stronger as q grows: $\partial \sigma^q / \partial q < 0$ and $\partial \sigma_+^q / \partial q < 0$. The belief σ_- does not depend on q, because negative signals must be genuine.

We also make the following assumptions:

Assumption 1. $C_a - \frac{c}{p} \in (\sigma^G, \sigma_+^G]$,

Assumption 2. $C_a - B \in (\sigma_-, \sigma_+^G]$.

These assumptions imply that the information in the report is potentially decision-relevant

---

[9]Disclosure is always truthful, in the sense that a government which discloses a possibly forged report $\mathbf{s}^q$ has not observed the genuine signals. Moreover, the public does not explicitly penalize the government for manipulating information, since the public is only concerned with how protest affects its own payoff.

[10]Simple computations yield $\sigma_+^q = \frac{\theta^2 + q^2(1-\theta)^2 + qR}{V + 2qR + q^2 V}$, $\sigma^q = \frac{(R/2) + q(1-\theta)^2}{R + qV}$ and $\sigma_- = \frac{(1-\theta)^2}{V}$, where $R \equiv 2\theta(1-\theta)$ and $V \equiv \theta^2 + (1-\theta)^2$.

8

for both players. Assumption 1 states that the public will protest if the government selects implementation when the evidence is mixed, but will accept implementation following two positive, genuine signals. Assumption 1 requires that c not be too large. If c was very large, the public would never protest and the analysis would be uninteresting.

Assumption 2 states that, even without the disciplining effect of public opinion ($E = 0$), the government prefers the status quo over implementation when both signals are negative, and prefers implementation over the status quo when two genuine signals are positive. This assumption ensures that the government always bears a cost for manipulating information because there are states when it prefers the status quo.

Together, Assumptions 1 and 2 imply that the government and the public always agree when the evidence is clear cut. They both favor implementation when $(\alpha, \alpha)^G$ and they both favor the status quo when $(\varnothing, \varnothing)^G$. Nevertheless, disagreement can arise when the signals are not genuine, or when genuine signals are mixed. In the latter case, the public prefers the status quo (by Assumption 1), but the government may prefer implementation (if B is sufficiently large). This disagreement is the source of the agency problem in our setting.

Observability of the Agency's Bias. An important issue is whether the public can observe the agency's bias q. For most of the paper, we will focus on the polar opposite scenarios of 'transparency' and 'secrecy'. Under transparency, both the agency's report and its levels of bias are observable, while under secrecy, neither is observable. Transparency should therefore be interpreted as an environment where information is easily accessible, not just about the contents of the report, but also about the staffing, track record and likely bias of the agency that drafts it. In contrast, under secrecy, information about the agency as well as the report is tightly guarded. Focusing on transparency versus secrecy allows us to keep the analysis tractable. Intermediate cases between transparency and secrecy are discussed in Section 6.

3 Transparency

We assume throughout this section that all information must be truthfully disclosed and that the agency is nonindependent. Since the government and the public both observe the signals $\mathbf{s}^q$ and the agency's bias q, they will share the same posterior beliefs. Let $\pi(p(\mathbf{s}^q), v(\mathbf{s}^q))$ denote the government's payoff given $\mathbf{s}^q$, where $p(\mathbf{s}^q)$ denotes the policy decision of the government

and $v(\mathbf{s}^q)$ denotes the public's decision whether to protest or accept implementation. Let $\Pr(\mathbf{s}^q | \mathbf{s}^G)$ be the probability of observing $\mathbf{s}^q$ conditional on genuine signals $\mathbf{s}^G$. For any given $q \in [0, 1]$, the government's expected payoff is

$$E(\pi^q) = \sum_{\mathbf{s}^G \in \{\alpha, \varnothing\}^2} \left[\sum_{\mathbf{s}^q \in \{\alpha, \varnothing\}^2} \pi(p(\mathbf{s}^q), v(\mathbf{s}^q) | \mathbf{s}^G) \Pr(\mathbf{s}^q | \mathbf{s}^G) \right] \Pr\left(\mathbf{s}^G\right). \tag{4}$$

Equation (4) shows that manipulating information affects the government through two distinct channels. Higher levels of bias undermine the government's ability to tailor its policy decision $p(\mathbf{s}^q)$ to the true state. This harms the government because its optimal decision is state-dependent. However, changing the distribution of observed signals $\Pr(\mathbf{s}^q | \mathbf{s}^G)$ also allows the government to shape public opinion, $v(\mathbf{s}^q)$. This can benefit the government by helping convince the public to accept implementation.

We begin with a preliminary result showing that the equilibrium level of bias is bounded from above.

Lemma 1. *The government will never choose a level of bias that always leads the public to protest implementation. Formally, in equilibrium, $q \in [0, q^{\max}]$, where $q^{\max} \in [0, 1)$ solves $\sigma_+^{q^{\max}} = C_a - \frac{c}{p}$.*

Because the public is rational, the weight it places on a positive report is decreasing in the level of bias. When $q > q^{\max}$, the bias is so large that the public disregards the report: citizens protest even when both signals are positive. The government strictly prefers setting a lower value of $q \in [0, q^{\max}]$, which provides better information for decision making and can also generate support for implementation through a positive report.

Having bounded from above the level of bias that can be optimal, we now examine which policy decisions are taken and supported in equilibrium. We begin with a partial result that simplifies the government's optimization problem. A full characterization of equilibrium play is provided later in Proposition 1.

Lemma 2. *In equilibrium, the government selects implementation when both signals are positive and the status quo when both signals are negative. The public accepts implementation if and only if both signals are positive.*

Lemma 2 easily follows from Assumptions 1 and 2 and the fact that in equilibrium $q \leq q^{\max}$. It shows that the public and the government always agree on the appropriate course

10

of action when the evidence is clear-cut (i.e., when the signals are both positive or both negative). What Lemma 2 does not show is whether the government will implement the policy when the evidence is mixed, in spite of public protests. To distinguish between the two relevant cases, we make the following definition.

Definition (discipline). Fix $q \leq q^{\mathrm{max}}$. The government is said to be *disciplined by public opinion* if it selects implementation if and only if both signals are positive.

A government that is disciplined by public opinion selects the status quo when the evidence is mixed, and so always enjoys public support. It always takes the public's preferred action and never faces protests.[11] For given $q \leq q^{\mathrm{max}}$, let $E(\pi_d^q)$ denote the government's payoff under discipline. Specifically, let $E(\pi_d^q)$ be a special case of (4) where (i) $q \leq q^{\mathrm{max}}$, (ii) the government selects implementation if and only if both signals are positive, and (iii) the public accepts implementation if and only if both signals are positive. Note that $E(\pi_d^q)$ incorporates all the requirements in Lemmas 1 and 2 as well as the notion of discipline.

A government that is not disciplined by public opinion will select implementation when the evidence is mixed, despite the public's protests. Let $E(\pi_{nd}^q)$ be the government's payoff in that case. Thus, $E(\pi_{nd}^q)$ is a special case of (4), where (i) $q \leq q^{\mathrm{max}}$, (ii) the government selects implementation if and only if the signals are positive or mixed, and (iii) the public accepts implementation if and only if both signals are positive. Both $E(\pi_d^q)$ and $E(\pi_{nd}^q)$ are explicitly computed in Appendix A.

To simplify the exposition of the results, we rule out corner solutions that would arise when the constraint $q \leq q^{\mathrm{max}}$ binds. Specifically, we assume

$$q^* \equiv \arg \max_{q \in [0,1]} E(\pi_d^q) < q^{\mathrm{max}}. \tag{5}$$

Like Assumption 2, condition (5) requires that B not to be too large. An explicit condition is provided in the appendix (see the proof of Proposition 1).

We can now state this section's main result.

Proposition 1. *Fix the value of $p \in (0,1)$, and let $E = L/(1-p)$ from (3). Then in the equilibrium of the transparency game, the public accepts implementation if and only if $\mathbf{s}^q = (\alpha, \alpha)$. Moreover*

---

[11]This is true because $q \leq q^{\mathrm{max}}$. If $q > q^{\mathrm{max}}$, then the public would always protest following implementation.

i. *If $B \leq C_a - \frac{1}{2}$, then the government selects implementation if and only if $\mathbf{s}^q = (\alpha, \alpha)$, and the equilibrium level of bias is zero.*

ii. *If $B \in (C_a - \frac{1}{2}, C_a - \frac{1}{2} + E)$, then the government selects implementation if and only if $\mathbf{s}^q = (\alpha, \alpha)$, and the equilibrium level of bias is $q^* = \frac{R}{V} \frac{B - C_a + 1/2}{C_a - B - \sigma_-} < q^{\max}$, where $R \equiv 2\theta(1 - \theta)$ and $V \equiv \theta^2 + (1 - \theta)^2$.*

iii. *If $B \geq C_a - \frac{1}{2} + E$, then there are two possible cases. In the first case, the government selects implementation if and only if $\mathbf{s}^q = (\alpha, \alpha)$, and the level of bias is $q^* = \frac{R}{V} \frac{B - C_a + 1/2}{C_a - B - \sigma_-} < q^{\max}$. In the second case, the government selects implementation if and only if $\mathbf{s}^q \neq (\varnothing, \varnothing)$, and the level of bias is $q^L(p, E) < q^*$, with $q^L(p, E) = 0$ if $p \leq 1/2$. The second case arises if and only if $E(\pi_d^{q^*}) \leq E(\pi_{nd}^{q^L(p,E)})$.*

Proposition 1 fully characterizes equilibrium play in the transparency game.[12] Case (i) deals with a situation where B is so small that, even in the absence of protests, the government would select the status quo when the evidence was mixed.[13] The government's interests are aligned with those of the public, so there is no need to manipulate information.

Case (ii) deals with a situation where the government would select implementation if $E = 0$ and the genuine signals were mixed, $(B > C_a - \frac{1}{2})$, but where public pressure leaves it unwilling to make an unpopular decision $(B < C_a - \frac{1}{2} + E)$. There is a conflict of interest between the government and the public, but the government still takes the public's preferred action so as to avoid protests.

However, precisely because public opinion is so powerful, the government now has an incentive to shape it. Note that the government's choice of bias affects the distribution of the observed signals $\mathbf{s}^q$. This has two effects on the government's payoff $E(\pi_d^q)$. On the one hand, higher levels of bias q reduce the quality of information available for decision making. Specifically, with probability $\frac{1}{2}q^2 V$, two genuine negative signals are transformed into two positive signals. This will result in the policy being implemented and in an expected loss for the government of $C_a - B - \sigma_-$, relative to the counterfactual where $q = 0$. On the other hand, the government wants to 'trick' the public into supporting implementation when the

---

[12]This equilibrium is unique given our particular tie-breaking rules that specify how players behave when indifferent. Different tie-breaking rules would only affect play in a region of parameter space (B, E) that has zero area, specifically for parameters such that $E(\pi_d^{q^*}) = E(\pi_{nd}^{q^L(p,E)})$.

[13]This follows from equation (2) and the fact that $\sigma_{Gov} \leq \frac{1}{2}$ when the evidence is mixed.

genuine signals are mixed. Manipulations help the government because they can transform mixed signals into two positive signals. This occurs with probability qR and yields a net benefit of $B - C_a + \frac{1}{2}$ to the government, relative to the counterfactual where $q = 0$. The optimal q balances precisely these gains from manipulation against the costs associated with poor decision making.[14]

The third case is when $B \geq C_a - \frac{1}{2} + E$. This case is more complicated because the government may or may not select implementation when the evidence is mixed. Condition $B \geq C_a - \frac{1}{2} + E$ implies that in the absence of bias, $q = 0$, the government would select implementation (no discipline). However, q is endogenous, and the government's optimal level of bias may differ from zero. The crucial observation here is that for any given signal realization $\mathbf{s}^q$, the government's incentive to select implementation is (weakly) decreasing in q. As q becomes large, observed α signals are more likely to be fake, leaving the government more reluctant to implement the policy. Setting a large level of bias may lead the government to select the status quo when the evidence is mixed (discipline).

We distinguish between two cases, depending on whether the optimal q is above or below a cutoff $\hat{q}$. The cutoff is defined so that, if $q \leq \hat{q}$, the government selects implementation after observing a mixed report (no discipline).[15] Thus, the government payoff on $[0, \hat{q}]$ is $E(\pi_{nd}^q)$. Conversely, if $q > \hat{q}$, then the government selects the status quo after observing a mixed report (discipline). Thus, the government payoff on $q \in (\hat{q}, q^{\max}]$ is $E(\pi_d^q)$.

Taken together, we obtain the following expression for the government's payoff as a function of q when $B \geq C_a - \frac{1}{2} + E$:

$$
E(\pi_{LB}^q) \equiv
\begin{cases}
E(\pi_{nd}^q) & \text{if } q \in [0, \hat{q}] \\
E(\pi_d^q) & \text{if } q \in (\hat{q}, q^{\max}]
\end{cases}
. \tag{6}
$$

When private benefits B are large (LB), the government will maximize $E(\pi_{LB}^q)$ with respect to $q \in [0, q^{\max}]$. Two types of equilibria can arise: one characterized by a low bias in information and by no discipline (when the optimal q lies on the interval $[0, \hat{q}]$), and another characterized by a large bias in information and by discipline (when the optimal q lies on the interval $(\hat{q}, q^{\max}]$), as in Case (ii).

---

[14]Manipulations can also transform two negative signals into mixed signals. Under discipline, however, this change is inconsequential because the government selects the status quo in both cases.

[15]Formally, $\hat{q}$ is implicitly defined by $\sigma^{\hat{q}} = C_a - B + E$. See Appendix B for an explicit characterization.

In the proof of Proposition 1, we show that the government's payoff is concave over $(\hat{q}, q^{\max}]$, and that it is convex over $[0, \hat{q}]$ whenever p is below a threshold that exceeds $1/2$. We give an explicit expression for the threshold in the proof of Proposition 2. This nonconcavity leads to an optimal level of the bias of either zero or $q^* \equiv \arg\max_{q \in [0,1]} E(\pi_d^q)$, as stated in Proposition 1(iii). For values of p above the threshold, the government's payoff is concave over $[0, \hat{q}]$ as well. The equilibrium level of bias when the government chooses no discipline may then be non-zero, $q^L(p, E)$, but it still remains below q^*, the equilibrium bias under discipline. We can say more precisely for which parameter values the government chooses discipline and high bias, $E(\pi_d^{q^*}) > E(\pi_{nd}^{q^L(p,E)})$, and for which parameter values it chooses no discipline and low bias, $E(\pi_d^{q^*}) \leq E(\pi_{nd}^{q^L(p,E)})$. However, for ease of exposition, we postpone this discussion until the following section.

Proposition 1 shows that for the most part, government behavior only depends on p and L through the expression $E = p/(1 - L)$. To distinguish between the different cases, what matters is the effectiveness of the public at constraining government actions. Whether the strength of public pressure arises primarily through the ability to block implementation (p close to 1) or to make the government suffer a high explicit loss (L large) is unimportant. However, the government's optimal behavior within case (iii) will depend independently on both p and L. Both the optimal level of bias without discipline, $q^L(p, E)$, and the condition $E(\pi_d^{q^*}) \leq E(\pi_{nd}^{q^L(p,E)})$, which determines whether or not the government chooses discipline, depend on p and L in ways other than the ratio $E = p/(1 - L)$. Put another way, holding fixed E at a value satisfying case (iii), the government's behavior may differ depending on whether p is high and L is low, or whether p is low and L is high. The next section addresses this point in more detail.

3.1 Can the Public Benefit from Biased Information?

Proposition 1 shows that, depending on parameter values, two different equilibrium outcomes can arise: one with no discipline and low bias ($q = q^L(p, E)$), and another with discipline and high bias $\left(q = q^* > q^L(p, E)\right)$. Moreover, under Case (iii), the government may be willing to choose discipline precisely because it is also able to manipulate information. If the bias were forced to be zero, so that the agency was independent, then the government would become less cautious and select implementation when the evidence was mixed.

This subsection explores the idea that, due to the positive equilibrium association between

bias and discipline, the government's ability to manipulate information may benefit the public. We begin with an illustrative example, describing the region of parameter space (B, E) for which this is indeed the case.

Figure 1 shows the equilibrium level of discipline and bias implied by Proposition 1, as a function of B and E, when $C_a = 0.6$, $\theta = 0.8$, and $c = 0$, in the limit as p tends to zero. The different values of E in the figure correspond to different values of $L = (1 - p)E$, where p is held fixed.
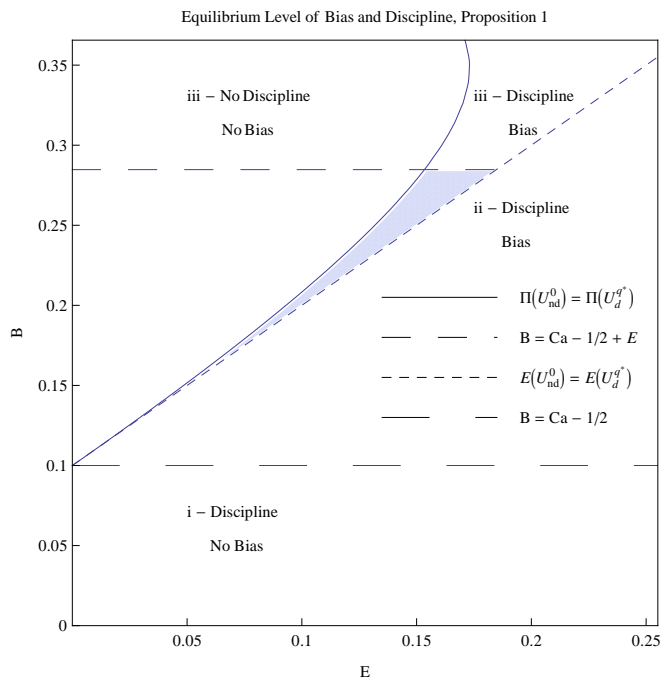

Figure 1: Equilibrium Level of Discipline and Bias, for $C_a = 0.6$, $\theta = 0.8$, $c = 0$, $\lim_{p \to 0}$

Figure 1 illustrates the regions of parameter space corresponding to the various cases of Proposition 1. Below the horizontal line $B = C_a - 1/2$, Case (i) implies the government will be disciplined by public opinion and will choose zero bias. Above this horizontal line but below the 45 degree line $B = C_a - 1/2 + E$, Case (ii) implies the government will be disciplined by public opinion, choose high bias q^*, and would remain disciplined even if it were forced to set a bias of zero. Above the 45 degree line are the two subregions corresponding to Case (iii), in a setting where the public's ability to block implementation through protest is low: one where the government chooses no discipline and no bias, and another where it chooses discipline and high bias q^*.

Figure 1 also shows that the boundary between the two subregions follows a threshold structure. For any given $B > C_a - 1/2$, the government chooses no discipline and low bias, $E(\pi_d^{q^*}) \leq E(\pi_{nd}^{q^L(p,E)})$, if and only if E is below a certain threshold. Intuitively, the government's incentive to choose discipline is increasing in the strength of public pressure, since discipline allows it to avoid protests. For all $B > C_a - 1/2$, this threshold is strictly greater than zero and strictly less than $B - C_a + 1/2$. This means that both subregions have strictly positive area.

In the subregion of Case (iii) with discipline and high bias, the government's ability to manipulate information may benefit the public, which is the case in the shaded area. Within this subregion, the government's ability to manipulate information has two effects on citizen welfare. Bias makes positive signals less reliable, so that the government selects the status quo when the evidence is mixed. This moderating effect of bias helps the public by making the government more cautious. However, bias also means that seemingly positive signals may be forgeries, which hurts the public by unduly stacking the deck in favour of implementation. The public will benefit from the government's ability to manipulate information if the gains from discipline generated by this moderating effect outweigh the losses due to biased decision making. Moreover, these losses are increasing in the level of bias. Because q^* is increasing in B and independent of E, there is a threshold value of B below which the public is willing to accept bias q^* to more closely align the government's interests with its own.[16] The public benefits from the government's ability to manipulate information in the shaded area of Figure 1, which is the part of subregion (iii - Discipline and Bias) where q^* is not excessively high.

More generally, as shown in Proposition 2 below, there is always a region of parameter space (B, E) corresponding to the shaded area in Figure 1, where the public strictly benefits from the government's ability to manipulate information.

Proposition 2. *Fix the value of $p \in (0, 1)$, and let $E = L/(1-p)$ from (3). Then for any C_a and any $\theta \in (\frac{1}{2}, 1)$, there are values of B and E such that a commitment not to manipulate information strictly hurts the public. Specifically, there exists $\overline{B} > C_a - \frac{1}{2}$, and $\underline{E}(B) < B - C_a + \frac{1}{2}$ for any $B \in (C_a - \frac{1}{2}, \overline{B})$, such that $E(U_d^{q^*}) > E(U_{nd}^{q^L(p,E)})$ if and only if $B \times E \in (C_a - \frac{1}{2}, \overline{B}) \times (\underline{E}(B), B - C_a + \frac{1}{2}]$. The region of parameter space (B, E) for which $E(U_d^{q^*}) > E(U_{nd}^{q^L(p,E)})$ is increasing in size with p and c, where $\frac{\partial}{\partial p}\underline{E}(B) < 0$, $\lim_{p \to 1} \underline{E}(B) = 0$,*

---

[16]The threshold value of B is identified by $E(U_d^{q^*}) = E(U_{nd}^G)$ when $q^* < q^{max}$ holds at this threshold, where B only affects the public's payoff via $q^* = \frac{R}{V} \frac{B - C_a + 1/2}{C_a - B - \sigma_-}$.

and $\frac{\partial}{\partial c}\overline{B} > 0.$

Proposition 2 confirms that the results from Figure 1 hold more generally. There is always a region of parameter space for which manipulations help the public, comprising exactly those pairs (B, E) for which B is sufficiently close to $C_a - \frac{1}{2}$, and for which E exceeds a threshold between zero and and $B - C_a + 1/2$. Proposition 2 also generates a number of new insights. First, it shows that manipulating information is more likely to help the public when there is a relatively large cost of protesting. As long as Assumption 1 holds, forcing the bias to zero will lead to protests, when the government implements based on evidence that is mixed. An increase in c hurts the public in such situations. Citizens are therefore willing to accept a higher level of bias in order to ensure the government remains disciplined by public opinion, leading to a higher cutoff $\overline{B}$.[17]

Second, and perhaps more subtly, Proposition 2 shows that manipulating information is more likely to help citizens when public pressure arises mainly from potentially blocking implementation (p large and L small), rather than imposing a high explicit loss on the government (L large and p small). Figure 1, where p is small and $c = 0$, therefore understates the size of parameter space where manipulations help the public. The intuition is that the government's payoff from choosing discipline is independent of p and L, since the public never protests. In contrast, the government's payoff from no discipline is decreasing in p, holding $E = L/(1-p)$ constant. A government choosing no discipline is willing to suffer the explicit loss due to protest in order to implement the new policy when the evidence is mixed. On the margin, increasing the public's ability to block implementation then hurts the government to a large extent, outweighing the drop in the explicit loss from protest.

It follows that an increase in p and drop in L, holding E constant, leads the government to choose discipline and high bias q^* in a larger region of parameter space. Redrawing Figure 1 with p fixed at a higher value would therefore yield a larger shaded region, where the ability to manipulate helps the public. In the limit as p tends to 1 and L tends to zero, the cutoff $\underline{E}(B)$ tends to zero for all $B \in (C_a - 1/2, \overline{B})$. The condition $E(\pi_d^{q^*}) > E(\pi_{nd}^{q^L(p,E)})$ then always holds so the government chooses discipline and high bias in the entire region corresponding to Case (iii). The shaded area in Figure 1 then extends to the vertical axis. Paradoxically, for a given level of public pressure, a strong ability to block new policy often leads the government to set high bias, but forcing this bias to zero will often hurt the public.

---

[17]This result takes into account condition (5), $q^* < q^{\max}$. See the appendix for more details.

The above discussion is predicated on the assumption that p and L change so as to leave the strength of public pressure E constant. Proposition 2 also shows what happens when E changes. Bias can only moderate government policy if public pressure is relatively weak. When $E > B - C_a + \frac{1}{2}$, so when p and L are sufficiently large, the government is always disciplined by public opinion, regardless of the level of bias. Public pressure then suffices to ensure that the government caters to the public. Thus, from the public's point of view, manipulations simply stack the deck in favor of implementation, which decreases their payoff. The present model therefore suggests that the independence of government agencies should unambiguously benefit the public in mature democracies, where E is large. In contrast, in less mature democracies, where governments care about public opinion but are not fully responsive to it, nonindependence may sometimes be socially optimal.

4   Secrecy

The previous section studied the case where the government must truthfully disclose both the signal realizations and the level of bias. This section analyzes the polar opposite scenario of *secrecy*: the government commits not to disclose either $\mathbf{s}^q$ or q. The main complication that arises is that, as in Canes-Wrone et al. (2001) and Maskin and Tirole (2004), the government now has private information. As a result, the public's choice will in general depend on the policy decision of the government, which potentially conveys information.[18]

Despite this complication, we can characterize equilibrium play. Let $R = 2\theta(1-\theta)$ and $V = \theta^2 + (1-\theta)^2$. Furthermore, let $\hat{\sigma} = \frac{\theta^2 + R}{1+R} \in \left(\frac{1}{2}, 1\right)$ be the public's belief that the state is good when $q = 0$, implementation is selected by the government, and the government is disciplined by public opinion.

Proposition 3. *In the equilibrium of the secrecy game, the level of bias is zero. The government selects implementation when both signals are positive and selects the status quo when both signals are negative. Moreover, fixing the value of $p \in (0,1)$, and letting $E = L/(1-p)$ from (3), we have*

i. *If $B \le C_a - \frac{1}{2}$, then the government selects the status quo when the evidence is mixed and the public accepts implementation whenever implementation is selected.*

---

[18]On the other hand, the public's choice v cannot depend on the realization of the signals $\mathbf{s}^q$ or the bias q because they are unobservable.

ii. If $B \in (C_a - \frac{1}{2}, C_a - \frac{1}{2} + E)$ and $\hat{\sigma} < C_a - \frac{c}{p}$, then the government selects implementation with probability $\check{s} = \left(\frac{V}{2R}\right) \frac{\sigma_+^G - C_a + \frac{c}{p}}{C_a - \frac{c}{p} - 1/2} \in (0,1)$ when the evidence is mixed and the public accepts implementation with probability $1 - \frac{\frac{1}{2} - C_a + B}{p(\frac{1}{2} - C_a + B) + (1-p)E} \in (0,1)$ when implementation is selected.

iii. If $B \in (C_a - \frac{1}{2}, C_a - \frac{1}{2} + E)$ and $\hat{\sigma} \geq C_a - \frac{c}{p}$, then the government selects implementation when the evidence is mixed and the public accepts implementation whenever implementation is selected.

iv. If $B \geq C_a - \frac{1}{2} + E$ and $\hat{\sigma} < C_a - \frac{c}{p}$, then the government selects implementation when the evidence is mixed and the public protests implementation whenever implementation is selected.

v. If $B \geq C_a - \frac{1}{2} + E$ and $\hat{\sigma} \geq C_a - \frac{c}{p}$, then the government selects implementation when the evidence is mixed and the public accepts implementation whenever implementation is selected.

The equilibrium in Proposition 3 exhibits several intuitive features.[19] First, consistent with conventional wisdom, secrecy is shown to be effective at protecting the integrity of the decision-making process. The government has no incentive to set a positive bias because neither q nor the signal realizations are observed by the public. Increasing q simply reduces the quality of information available to the government, so the equilibrium level of bias is zero.

A second intuitive feature of the equilibrium is that, as the government's private benefits B grows large, the government is less likely to be disciplined by public opinion. Proposition 3 shows that the government always selects implementation when the signals are positive, and always selects the status quo when the signals are negative. Thus, for the government to be disciplined by public opinion, we only need to check whether the government selects the status quo when the signals are mixed. Proposition 3 shows that, when the signals are mixed, implementation is always selected when B is large (cases (iv)-(v)), and it is often selected when B is intermediate (cases (ii)-(iii)). It is only when B is small that the status quo is always selected (case (i)). Thus, as B grows large, the government is less likely to cater to public opinion.

---

[19]As in Proposition 1, this equilibrium is unique given our particular tie-breaking rules. Different tie-breaking rules would only affect play in a region of parameter space (B, E) that has zero area, specifically when $B = C_a - \frac{1}{2}$ and $B = C_a - \frac{1}{2} + E$ if $\hat{\sigma} < C_a - \frac{c}{p}$, and when $\hat{\sigma} = C_a - \frac{c}{p}$.

It is also instructive to compare the equilibrium outcomes under transparency and secrecy (Propositions 1 and 3). More cases must be distinguished under secrecy than under transparency (five versus three). Under secrecy, when the government selects implementation, the public cannot observe whether the signal are positive or mixed. The condition $\hat{\sigma} < C_a - c/p$ describes scenarios where mistaken implementation is sufficiently costly for the public to protest if implementation is selected. When mistaken implementation is less costly ($\hat{\sigma} \geq C_a - c/p$), the situation is reversed. These complications do not arise under transparency because the public can observe the signals, so that beliefs $\hat{\sigma}$ play no role.

Because public opinion is influenced by the government's policy decision, a mixed strategy equilibrium can arise under secrecy (case ii), which does not happen under transparency. In this case, if the government always implements when the evidence is mixed, then the public would always protest following implementation, by $\hat{\sigma} < C_a - \frac{c}{p}$. But given this protest, the government would actually prefer to select the status quo, by $B < C_a - \frac{1}{2} + E$. Similarly, if the government never implements when the evidence is mixed, then the public would never protest, leading the government to prefer implementation, by $B > C_a - \frac{1}{2}$. It follows that in equilibrium, the government and public must play mixed strategies, where the probability of implementation after mixed evidence leaves the public indifferent about protesting following implementation, and where the probability of protesting following implementation leaves the government indifferent about implementing when the evidence is mixed.

Finally, from the public's point of view, the choice between transparency and secrecy involves a key trade-off between manipulations and discipline. Manipulations are always (weakly) lower under secrecy, while discipline is always (weakly) higher under transparency. That manipulations are lower under secrecy is obvious as $q = 0$. Let us therefore compare transparency and secrecy in terms of discipline. When there is no conflict of interest (case (i)), the government is disciplined by public opinion under both scenarios. In contrast, when the conflict of interest is intermediate, the government is always disciplined by public opinion under transparency (Proposition 1(ii)) but not under secrecy (Proposition 3(ii)-(iii)). Furthermore, when the conflict of interest is large, the government is sometimes disciplined by public opinion under transparency (Proposition 1(iii)) but never under secrecy (Proposition 3(iv)-(v)). Thus discipline is always at least as likely under transparency as under secrecy.

This lack of discipline under secrecy is caused by a relative lack of accountability. Without observing the report, citizens cannot determine exactly why a particular decision was taken.

For example, the government's decision to select implementation could be based on strong evidence ($\mathbf{s}^G = (\alpha, \alpha)$) or mixed evidence ($\mathbf{s}^G = (\alpha, \varnothing)$). The public would only like to punish the government in the latter case but it cannot do so without seeing the report. As a result, the government is less accountable and thus less responsive to public desires. It is easy to construct examples where, because of this trade-off between manipulations and accountability, either transparency or secrecy is preferred by the public.

5 Independence and Optimal Constitutions

So far we have assumed that the government can easily interfere with the workings of the agency in charge of collecting information. This is a reasonable assumption if, as in the U.S., the President appoints and can remove the heads of the executive agencies, thus exerting enormous influence over their policy decisions. Sometimes, however, executive influence over government agencies is more limited. Of special interest is the case of *independent* agencies such as the Federal Trade Commission in the U.S. and the Bank of England in the U.K. These agencies are not subject to the same degree of political control as other executive agencies and are insulated from political pressure, for instance by limiting the removal of their heads to certain causes.

This section considers the implications of granting full independence to the government agencies in charge of collecting information. Formally, independence is modelled as a commitment not to manipulate information. Thus, an independent agency will carry out its job as objectively as possible. We first compare transparency and secrecy under the assumption that the agency is independent.

Proposition 4. *Suppose the government agency is independent (i.e., $q \equiv 0$). Then the public's payoff is always higher under transparency than under secrecy.*

The intuition for this result is straightforward: if information cannot be manipulated, then only accountability matters, and transparency is in the interests of the public.

Having established this benchmark result, we now consider the more interesting case where both the disclosure rule (transparency or secrecy) and the degree of insulation of the government agency (independence or nonindependence) can be chosen to maximize the public's welfare. Following previous work, we refer to the stage when society decides the

rules of the game as the 'constitutional' stage. The four constitutions we consider are shown in Table I.

Table I: Constitutions

I	Transparency & Independent Agency	II	Transparency & Nonindependent Agency
III	Secrecy & Independent Agency	IV	Secrecy & Nonindependent Agency

A constitution is said to be optimal if it maximizes the public's welfare. Proposition 5 characterizes optimal constitutions.

Proposition 5. *An optimal constitution always involves transparency. The comparison between Constitution I (transparency & independent agency) and Constitution II (transparency & nonindependent agency) is ambiguous.*

In an environment where information disclosure creates incentives for manipulation, it is perhaps surprising that transparency is always optimal. The intuition for this result is simple: the chief advantage of secrecy – unbiased information – can more effectively be achieved by insulating the government agency from political pressure. To see this more formally, note that by Proposition 4, Constitution I (transparency & independent agency) dominates Constitution III (secrecy & independent agency). Moreover, the two constitutions involving secrecy (Constitutions III and IV) are payoff equivalent because under secrecy q is always equal to zero in equilibrium. Thus transparency (Constitution I) always dominates secrecy (Constitutions III and IV).[20]

However, this result does not imply that granting independence to government agencies is necessarily in the public interest. Biased information can induce the government to behave more cautiously, thus mitigating the agency conflict between the government and the public (see Propositions 1 and 2). As a result, the comparison between Constitution I and Constitution II is ambiguous. Constitution II (non-independence) can be optimal if the conflict of interest between the government and the public is not too large, so that the benefits of discipline outweigh the costs of the agencies' small equilibrium bias.[21]

---

[20]This result requires the combination of transparency and independence to be available at the constitutional stage. If independence was not feasible, the trade-off between manipulations and accountability highlighted in the previous section would obviously reappear.

[21]If the conflict of interest is large (large B), then equilibrium bias can be large as well, so that Constitution

22

6   Extensions

The analysis so far has focused on the polar opposite scenarios of transparency and secrecy. Under transparency both the report and the bias in information are observable, while under secrecy neither is observable. This section briefly discusses two intermediate scenarios, disclosure of the report with unobservable bias and nondisclosure of the report with observable bias, as well as a third scenario where disclosure is voluntary. It also touches on possible alternative assumptions about public protest. More details about the results in this section are available from the authors upon request.

Disclosure with Unobservable Bias. We begin with the case when the report is disclosed but the public does not observe the bias of the agency. It can be shown that if assumption (5) holds, then the equilibrium outcome when q is unobservable is exactly the same as when q is observable. Thus, under assumption (5), Proposition 1 is not affected by the unobservability of q.

The idea behind this result is simple. When q is not observable, the public must form some conjecture about the level of bias chosen by the government. In equilibrium, this conjecture must be correct. Assumption (5) ensures that, when the public believes that $q < q^{\max}$, then the government will actually choose a level of bias $q < q^{\max}$. Thus the public's belief can be made consistent with the play of the game. In particular, the government's incentives to set any $q < q^{\max}$ are just as in Section 3, so the same equilibrium as in Proposition 1 (with observable q) can be supported.[22]

Nondisclosure with Observable Bias. An alternative scenario arises when the government commits not to disclose the report but the agency bias is observable. This scenario is plausible if the public is well-informed about the reputation and policy dispositions of the individuals working for the agency, even though the report is not disclosed.

Compared to the case where bias is unobservable (secrecy), an interesting new effect can arise. Specifically, the government may choose a strictly positive bias to commit itself to a more congruent decision rule. Intuitively, by appointing a head of the agency who is well-

---

I (independence) will often be optimal. In this situation, if independence is not feasible, then the public may actually benefit by reducing its influence on the government (reducing E) under Constitution II, sacrificing discipline in order to eliminate all bias.

[22]If assumption (5) does not hold, then equilibrium behavior is more complicated and will generally involve randomization over q.

known to be biased in favor of implementation, the government can credibly commit not to select implementation when the undisclosed signals are mixed. A mixed report from a very biased bureaucrat provides very little evidence in support of implementation.

The optimal choice of q is therefore determined by two conflicting effects. On the one hand, manipulations reduce the quality of information available to the government, which reduces its payoff. On the other hand, a sufficiently high level of bias allows the government to commit to a decision rule that the public prefers. This can induce the public to support the government policy more often, thereby increasing the government payoff.

That being said, this new effect of nondisclosure with observable bias is only present if the government needs to convince the public to support implementation. Proposition 3 shows that the public often supports the government's decision to implement even if the agency is unbiased ($q = 0$) (see cases (i), (iii) and (v)). In this sense, these elements of Proposition 3 will continue to hold whether or not q is observable.

Voluntary Disclosure. We have assumed so far that the government must either disclose the contents of the report or must keep it secret. More commonly, however, policymakers have discretion as to whether to release information. We now consider a variant of the model where the government cannot commit to any disclosure rule: disclosure is voluntary. We argue that voluntary disclosure will effectively result in all information being disclosed. Indeed, since information is hard in this model, Milgrom's (1981) 'unraveling' result applies.

To see the logic of this result, suppose the public expects the government to disclose favorable information (the α signals). Thus nondisclosure is interpreted as evidence that the information is unfavorable (a $\varnothing$ signal), which provides the government with a strong incentive to disclose favorable information. Specifically, a government that receives two positive signals will disclose them and implement the policy with public support. When the evidence is mixed or unfavorable, whether or not the signals are disclosed is inconsequential because the public would realize that at least one of them is unfavorable. Thus they would not support implementation. All information is therefore revealed, and the analysis would proceed as in Section 3.

Alternative Assumptions about Public Protest. We have assumed that the public protests implementation if and only if doing so increases its expected payoff. This assumption of rationality limits the extent to which the public can influence government policy, by

ruling out protest when the government chooses the public's preferred action (implementing following a positive report). If the public could commit to a punishment of magnitude B whenever the government selected implementation, then it could resolve the agency problem in our setting. For any bias q, interests would be fully aligned, and the government would be disciplined by public opinion. The government would then find it optimal to set $q = 0$ since the punishment it faces depends only on its actions, not on the content of the report. The assumption of a rational public rules out such punishments, which would be highly effective, but which are simply not credible.

Our analysis does show what occurs when the public has the strongest influence possible, given its inability to commit to a non-credible punishment. This can either reflect a situation where L is very large or where p is close to 1. Looking back at Figure 1, the first case corresponds to a large value of E in Region (ii), and the second case corresponds to a situation where $E(\pi_d^{q^*}) > E(\pi_{nd}^{q^L(p,E)})$ in almost all of Region (iii). Either way, the government will be disciplined by public opinion and set $q^* > 0$. This positive bias shows that the agency problem our analysis identifies will persist. Moreover, when p is close to 1, Proposition 2 implies that forcing the bias to zero will hurt the public in a large part of Region (iii).

Our conclusions also continue to hold if the public's decision to protest is continuous rather than binary. Suppose that the public sets protest intensity $e \in [0, 1]$, and pays increasing and convex costs, $c(e)$, with $c(0) = c'(0) = 0$. Protest then blocks implementation with probability ep and makes the government suffer an explicit loss of eL. Our previous analysis effectively assumes that $e \in \{0, 1\}$ and $c \equiv c(1)$. Allowing for continuous protest intensity leads to only minor differences: in Proposition 1, the protest intensity under no discipline will now be increasing in bias $q^L(p, E)$, and in the mixed strategy equilibrium of Proposition 3 (ii), the public will set some $e \in (0, 1)$ with certainly to make the government indifferent about implementation. However, all our main results will remain unchanged.

7   Conclusion

This paper develops a model where disclosure of information gives the government an incentive to "fix the evidence" around its ex ante favored policy. Decision-relevant information is collected by an agency, but the government can distort this process, for instance by staffing the agency with biased individuals. The key trade-off the government faces is between protecting the quality of the information available for public decision making (if the agency is

unbiased) and molding public opinion (if the agency is biased). Surprisingly, we find that insulating the agency from political pressure, so that the agency is always unbiased, is not necessarily in the public interest. A biased information gathering process can in fact induce the government to act more cautiously in response to information supporting its ex ante preferred policy. This moderating effect of bias can more than outweigh the welfare losses caused by biased information.

We are not the first to study whether government agencies should be insulated from external or political pressures. Moe (1989, 1990) argues that government agencies are sometimes intentionally created to be unresponsive to political pressures to alleviate the risk of political power fluctuations. Prendergast (2003) points out that bureaucrats' tendency to inefficiently accede to customer demands may require appropriate organizational responses, such as insulating government agencies from customer complaints. Betts (2004) notes that a close connection between the President and top intelligence officials may be preferable to the lack of such a connection because the risks of insulation and unresponsiveness often far outweigh those of politicization. This paper highlights a novel drawback of bureaucratic independence: the risk that candid advice from government agencies may make policymakers very responsive to information supporting their ex ante favored policy, thus exacerbating the conflict of interest between the government and the public.

It remains an open question as to exactly which of our conclusions extend to a setting where the government's policy preference is private information. A government that is relatively new on the political scene, facing a public unsure about the size of the conflict of interest, may try to convey information through its policy choice. If there is no bureaucratic independence, then it may even try to convey information through its choice of agency bias. We would expect the main feature of our analysis, that biased information can make the government more cautious, to apply in such a setting. There may also be additional insights as how agency bias may help or hinder the government to credibly signal its type. We view this as an interesting avenue for future research.

are our own.

References

Alesina, A., Tabellini, G. 2007. Bureaucrats or politicians? Part I: a single policy task. *American Economic Review* 97: 169-179.

Alesina, A., Tabellini, G. 2008. Bureaucrats or politicians? Part II: multiple policy tasks. *Journal of Public Economics* 92: 426-447.

Ashworth, S. 2012. Electoral accountability: Recent theoretical and empirical work. *Annual Review of Political Science* 15: 183-201

Ashworth, S., Shotts, K.W. 2010. Does informative media commentary reduce politicians' incentives to pander? *Journal of Public Economics* 94: 838-847.

Banisar, D. 2004. The Freedominfo.org Global Survey. Freedom of Information and Access to Government Record Laws around the World. Available at http://www.freedo-minfo.org/documents/global_

Besley, T., Prat, A. 2006. Handcuffs for the grabbing hand? Media capture and political accountability. *American Economic Review*, vol. 96, pp. 720–736.

Betts, R.K. 2004. The new politics of intelligence. *Foreign Affairs* 83: 2-8.

Canes-Wrone, B., Herron, M.C., Shotts, K.W. 2001. Leadership and pandering: A theory of executive policymaking. *American Journal of Political Science* 45: 532-550.

Committee of Privy Counsellors. Chairman: The Rt Hon The Lord Butler of Brockwell KG GCB CVO. 2004. Review of intelligence on weapons of mass destruction. Downloadable at http://www.butlerreview.org.uk/

Cremer, J. 1995. Arm's length relationships. *Quarterly Journal of Economics* 110: 275-295.

Daley, B., Snowberg E. 2011. Even if its not bribery: The case for campaign finance reform. *Journal of Law, Economics, and Organization* 27: 324-349.

Fox, J. 2007. Government transparency and policymaking. *Public Choice* 131: 23-44.

Fox, J., Jordan S.V. 2011. Delegation and accountability. *Journal of Politics* 73: 831-844.

Fox, J., Shotts K.W. 2009. Delegates or trustees? A theory of political accountability.

Journal of Politics 71: 1225-1237.

Gailmard, S., Patty, J.W. 2007. Slackers and zealots: Civil service, policy discretion, and bureaucratic expertise. *American Journal of Political Science* 51: 873–889.

Holmstrom, B. 1979. Moral hazard and observability. *Bell Journal of Economics* 10: 74-91.

Levy, G. 2007. Decision making in committees: Transparency, reputation and voting rules. *American Economic Review* 97: 150-168.

Marschak, J., Radner, R. 1972. *Economic Theory of Teams*, Yale University Press, New Haven and London.

Maskin, E., Tirole, J. 2004. The politician and the judge. *American Economic Review* 94: 1034-1051.

Milgrom, P. 1981. Good news and bad news: Representation theorems and applications. *Bell Journal of Economics* 12: 380-391.

Moe, T.M. 1989. The Politics of Bureaucratic Structure. In Chubb, J.E., Peterson P.E. (eds.), *Can the Government Govern?* Washington, D.C.: The Brookings Institution.

Moe, T.M. 1990. Political institutions: The neglected side of the story. *Journal of Law, Economics, & Organization* 6: 213-253.

Patty, J.W. 2009. The politics of biased information. *Journal of Politics* 71: 385-397.

Prat, A. 2005. The wrong kind of transparency. *American Economic Review* 104: 946-59.

Prat, A. 2006. The more closely we are watched, the better we behave? In Hood, C., Heald, D. (eds.), *Transparency: The Key to Better Governance?* Oxford University Press, Oxford.

Prendergast, C. 1993. A theory of "Yes Men". *American Economic Review* 83: 757-770.

Prendergast, C. 2003. The limits of bureaucratic efficiency. *Journal of Political Economy* 111: 929-958.

Prendergast, C. 2007. The motivation and bias of bureaucrats. *American Economic Review* 97: 180-196.

Roberts, A. 2006. Dashed expectations: Governmental adaptation to transparency rules. In Hood, C., Heald, D. (eds.), *Transparency: The Key to Better Governance?* Oxford University

Press, Oxford.

Shotts, K.W., Wiseman, A.E. 2010. The politics of investigations and regulatory enforcement by independent agents and cabinet appointees. *Journal of Politics* 72: 209-226.

Stasavage, D. 2004. Open-door or closed-door? Transparency in domestic and international bargaining. *International Organization* 58: 667-703.

Swank, J., Swank, O., Visser, B. 2008. How committees of experts deal with the outside world: Some theory, and evidence from the FOMC. *Journal of the European Economic Association* 6: 478-486.

Warren, P.L. 2012. Independent auditors, bias, and political agency. *Journal of Public Economics* 96: 78–88.

Ujhelyi, G. 2014. Civil service reform. *Journal of Public Economics* 118: 15-25.

Appendix A: Payoffs

This appendix derives explicit expressions for the government payoff and citizen welfare in the transparency case.

The Government Payoff. The government payoff is given by

$$E(\pi^q) = \sum_{\mathbf{s}^G \in \{\alpha, \varnothing\}^2} \left[\sum_{\mathbf{s}^q \in \{\alpha, \varnothing\}^2} \pi(p(\mathbf{s}^q), v(\mathbf{s}^q)|\mathbf{s}^G) \Pr(\mathbf{s}^q|\mathbf{s}^G) \right] \Pr\left(\mathbf{s}^G\right), \tag{A1}$$

where $\pi(p(\mathbf{s}^q), v(\mathbf{s}^q)|\mathbf{s}^G)$ is the government expected payoff when the observed signals are $\mathbf{s}^q$ and the genuine signals are $\mathbf{s}^G$, and q is the probability that a genuine $\varnothing$ signal is transformed into fake α signal.

To compute the probabilities in (A1), note that $\Pr((\alpha, \alpha)^q) = \Pr((\alpha, \alpha)^q \,|\, (\alpha, \alpha)^G) \Pr((\alpha, \alpha)^G) + 2\Pr((\alpha, \alpha)^q \,|\, (\alpha, \varnothing)^G) \Pr((\alpha, \varnothing)^G) + \Pr((\alpha, \alpha)^q \,|\, (\varnothing, \varnothing)^G) \Pr((\varnothing, \varnothing)^G)$ and so on. Moreover, $\Pr((\alpha, \alpha)^q \,|\, (\alpha, \alpha)^G) = 1$, $\Pr((\alpha, \alpha)^q \,|\, (\alpha, \varnothing)^G) = q$, $\Pr((\alpha, \alpha)^q \,|\, (\varnothing, \varnothing)^G) = q^2$ and so on. Finally, $\Pr((\alpha, \alpha)^G) = \Pr((\varnothing, \varnothing)^G) = \frac{1}{2}(\theta^2 + (1-\theta)^2)$, $\Pr((\alpha, \varnothing)^G) = \Pr((\varnothing, \alpha)^G) = \theta(1-\theta)$.

Because in equilibrium Lemmas 1 and 2 must hold, we compute $E(\pi^q)$ under the assumption that (i) the public accepts implementation if and only if both signals are positive, (ii) the government selects implementation when both signals are positive, and (iii) the government selects the status quo when both signals are negative. Thus, in equilibrium only two cases can emerge: either the government is disciplined by public opinion (thus it selects the status quo when the evidence is mixed) or the government is not disciplined by public opinion (thus it selects implementation when the evidence is mixed).

If the government is disciplined by public opinion, then (A1) is given by

$$\begin{aligned}
E(\pi_d^q) = {} & \frac{1}{2}V\pi(p = a, v = nd|(\alpha, \alpha)^G) + qR\pi(p = a, v = nd|(\alpha, \varnothing)^G) + \frac{1}{2}q^2 V\pi(p = a, v = nd|(\varnothing, \varnothing)^G) \\
& + (1-q)R\pi(p = n|(\alpha, \varnothing)^G) + q(1-q)V\pi(p = n|(\varnothing, \varnothing)^G) \\
& + \frac{1}{2}(1-q)^2 V\pi(p = n|(\varnothing, \varnothing)^G),
\end{aligned} \tag{A2}$$

where

$$V \equiv \theta^2 + (1-\theta)^2 = \Pr((\alpha, \alpha)^G) + \Pr((\varnothing, \varnothing)^G),$$
$$R \equiv 2\theta(1-\theta) = 2\Pr((\alpha, \varnothing)^G),$$

(the subscript d stands for discipline). It is simple to see that $\Pr((\alpha, \alpha)^q \,|\, (\alpha, \alpha)^G) \Pr((\alpha, \alpha)^G) = \frac{1}{2}V$, $2\Pr((\alpha, \alpha)^q \,|\, (\alpha, \varnothing)^G) \Pr((\alpha, \varnothing)^G) = qR$, and so forth. In terms of notation, the first line

of (A2) refers to situations where both signals are positive, the government implements, and the public accepts. The second and third lines refer respectively to situations where the evidence is mixed and where both signals are negative, and the government does not implement.

By contrast, if the government is not disciplined by public opinion, then (A1) is given by

$$
\begin{aligned}
E(\pi_{nd}^q) = \frac{1}{2}V\pi(p=a,v=nd|(\alpha,\alpha)^G) &+ qR\pi(p=a,v=nd|(\alpha,\varnothing)^G) + \frac{1}{2}q^2V\pi(p=a,v=nd|(\varnothing,\varnothing)^G) \\
&+ (1-q)R\pi(p=a,v=d|(\alpha,\varnothing)^G) + q(1-q)V\pi(p=a,v=d|(\varnothing,\varnothing)^G) \\
&+ \frac{1}{2}(1-q)^2V\pi(p=d|(\varnothing,\varnothing)^G),
\end{aligned}
\tag{A3}
$$

(the subscript nd stands for no discipline). Comparing with (A2), expression (A3) is identical to except for its second line, which refers to situations where the evidence is mixed, the government implements, and the public protests.

Computing the conditional payoffs π is straightforward. For instance, $\pi(p=a,v=nd|(\alpha,\alpha)^G) = -C_a\left(1-\sigma_+^G\right) + B = -C_a(1-\frac{\theta^2}{V}) + B$, $\pi(p=a,v=d|(\alpha,\varnothing)^G) = (1-p)(-C_a\left(1-\sigma^G\right)+B)+p(-C_n\sigma^G)-L = (1-p)(-\frac{1}{2}C_a+B)+p(-\frac{1}{2}C_n)-L$, $\pi(p=n|(\varnothing,\varnothing)^G) = -C_n\sigma_- = -\frac{(1-\theta)^2}{V}C_n$ and so on. Plugging these values into (A2) and (A3) yields

$$
\begin{aligned}
E(\pi_d^q) = \frac{1}{2}V\left[-C_a\left(1-\frac{\theta^2}{V}\right)+B\right] &+ qR\left[-\frac{1}{2}C_a+B\right] + \frac{1}{2}q^2V\left[-C_a\left(1-\frac{(1-\theta)^2}{V}\right)+B\right] \\
&+ (1-q)R\left[-\frac{1}{2}C_n\right] + q(1-q)V\left[-\frac{(1-\theta)^2}{V}C_n\right] \\
&+ \frac{1}{2}(1-q)^2V\left[-\frac{(1-\theta)^2}{V}C_n\right],
\end{aligned}
\tag{A4}
$$

and

$$
\begin{aligned}
E(\pi_{nd}^q) = \frac{1}{2}V\left[-C_a\left(1-\frac{\theta^2}{V}\right)+B\right] &+ qR\left[-\frac{1}{2}C_a+B\right] + \frac{1}{2}q^2V\left[-C_a\left(1-\frac{(1-\theta)^2}{V}\right)+B\right] \\
&+ (1-q)R\left[(1-p)\left(-\frac{1}{2}C_a+B\right)+p\left(-\frac{1}{2}C_n\right)-L\right] \\
&+ q(1-q)V\left[(1-p)\left(-C_a\left(1-\frac{(1-\theta)^2}{V}\right)+B\right)+p\left(-\frac{(1-\theta)^2}{V}C_n\right)-L\right] \\
&+ \frac{1}{2}(1-q)^2V\left[-\frac{(1-\theta)^2}{V}C_n\right].
\end{aligned}
\tag{A5}
$$

To simplify the computations in Appendix B, it is helpful to normalize $E(\pi_d^q)$ and $E(\pi_{nd}^q)$ by

subtracting $E(\pi_{nd}^G)$ from both. Since

$$E(\pi_{nd}^G) = \frac{1}{2}V\left[-C_a\left(1-\frac{\theta^2}{V}\right)+B\right]+R\left[(1-p)\left(-\frac{1}{2}C_a+B\right)+p\left(-\frac{1}{2}C_n\right)-L\right]+\frac{1}{2}V\left[-\frac{(1-\theta)^2}{V}C_n\right],$$
(A6)

(simply set $q=0$ in (A5)), we obtain

$$\Delta\pi_{d,nd}^q = RL - \frac{1}{2}q^2V\left[C_a - B - \sigma_-\right] + (1-p-q)R\left[C_a - \frac{1}{2} - B\right].$$
(A7)

$$\Delta\pi_{nd,nd}^q = qR\left[-p\left(C_a - \frac{1}{2} - B\right)+L\right]-\frac{1}{2}q^2V\left[C_a - B - \sigma_-\right]-(1-q)qV\left[(1-p)(C_a - B - \sigma_-)+L\right].$$
(A8)

where $\sigma_- = \frac{(1-\theta)^2}{V}$.

Citizen Welfare. Next, we derive the public's payoff (citizen welfare) in the transparency case. Citizen welfare is given by

$$E(U^q) = \sum_{\mathbf{s}^G\in\{\alpha,\varnothing\}^2}\left[\sum_{\mathbf{s}^q\in\{\alpha,\varnothing\}^2}U(p(\mathbf{s}^q),v(\mathbf{s}^q)|\mathbf{s}^G)\Pr(\mathbf{s}^q|\mathbf{s}^G)\right]\Pr\left(\mathbf{s}^G\right),$$
(A9)

where $U(p(\mathbf{s}^q),v(\mathbf{s}^q)|\mathbf{s}^G)$ denotes the public's payoff when the observed signals are $\mathbf{s}^q$ and the genuine signals are $\mathbf{s}^G$.

Because in equilibrium Lemmas 1 and 2 must hold, we also compute $E(U^q)$ under the assumption that (i) the public accepts implementation if and only if both signals are positive, (ii) the government selects implementation when both signals are positive, and (iii) the government selects the status quo when both signals are negative. Again, two cases can arise.

If the government is disciplined by public opinion, then (A9) becomes

$$E(U_d^q) = \frac{1}{2}V\left[-C_a\left(1-\frac{\theta^2}{V}\right)\right]+qR\left[-\frac{1}{2}C_a\right]+\frac{1}{2}q^2V\left[-C_a\left(1-\frac{(1-\theta)^2}{V}\right)\right] \quad (A10)$$
$$+(1-q)R\left[-\frac{1}{2}C_n\right]+q(1-q)V\left[-\frac{(1-\theta)^2}{V}C_n\right]$$
$$+\frac{1}{2}(1-q)^2V\left[-\frac{(1-\theta)^2}{V}C_n\right],$$

32

(as always, the subscript d stands for discipline).[23] Comparing (A10) with (A4) shows that $E(U_d^q)$ is equal to $E(\pi_d^q)$ but with the term B set to zero.

If the government is not disciplined by public opinion, then (A9) becomes

$$
\begin{aligned}
E(U_{nd}^q) = {} & \frac{1}{2}V\left[-C_a\left(1-\frac{\theta^2}{V}\right)\right] + qR\left[-\frac{1}{2}C_a\right] + \frac{1}{2}q^2V\left[-C_a\left(1-\frac{(1-\theta)^2}{V}\right)\right] \\
& + (1-q)R\left[(1-p)(-\frac{1}{2}C_a) + p(-\frac{1}{2}C_n) - c\right] \\
& + q(1-q)V\left[(1-p)(-C_a\left(1-\frac{(1-\theta)^2}{V}\right)) + p(-\frac{(1-\theta)^2}{V}C_n) - c\right] \\
& + \frac{1}{2}(1-q)^2V\left[-\frac{(1-\theta)^2}{V}C_n\right].
\end{aligned}
\tag{A11}
$$

(the subscript nd stands for no discipline). Similarly, comparing (A11) with (A5) shows that $E(U_{nd}^q)$ is equal to $E(\pi_{nd}^q)$ but with B set to zero and L replaced by c.

We also normalize (A10) and (A11) by subtracting $E(U_{nd}^G)$ from both. This yields

$$
\Delta U_{nd,nd}^q = qR\left[-p(C_a - \frac{1}{2}) + c\right] - \frac{1}{2}q^2V\left[C_a - \sigma_-\right] - (1-q)\,qV\left[(1-p)(C_a - \sigma_-) + c\right],
\tag{A12}
$$

$$
\Delta U_{d,nd}^q = Rc - \frac{1}{2}q^2V\left[C_a - \sigma_-\right] + (1-p-q)\,R\left[C_a - \frac{1}{2}\right].
\tag{A13}
$$

Appendix B: Proofs

Proof of Lemma 1. To prove Lemma 1, it suffices to show that setting any $q > q^{\max}$ is dominated by setting $q = 0$. The government payoff when $q = 0$ is the maximum between $E(\pi_d^G)$ and $E(\pi_{nd}^G)$. To derive the government payoff when some $q > q^{\max}$ is selected, we use two facts. First, if $q > q^{\max}$, then, by definition of $q^{\max}$, the public always protests implementation. Second, if $\mathbf{s}^q = (\varnothing, \varnothing)$, then by Assumption 2 the government must choose the status quo. Thus, when $q > q^{\max}$, only three cases must be considered.

Case (i): the government never selects implementation. This strategy obviously yields a lower payoff than $E(\pi_d^G)$. In both cases, the public never protests. In the latter case, however,

---

[23]The probabilities in (A9) have been computed above when deriving the government payoff. The payoffs conditional on the true underlying signals are also easy to derive. For instance, $U(p = a, v = nd|(\alpha,\alpha)^G) = -C_a\left(1-\sigma_+^G\right) = -C_a(1-\frac{\theta^2}{V})$, $U(p = a, v = d|(\alpha,\varnothing)^G) = (1-p)(-C_a\left(1-\sigma^G\right)) + p(-C_n\sigma^G) - c = (1-p)(-\frac{1}{2}C_a) + p(-\frac{1}{2}C_n) - c$, $U(p = n|(\varnothing,\varnothing)^G) = -C_n\sigma_- = -\frac{(1-\theta)^2}{V}C_n$ and so forth.

the government selects implementation when the signals are both positive. By Assumption 1, that yields a larger payoff than selecting the status quo.

Case (ii): the government selects implementation if and only if $\mathbf{s}^q = \{(\alpha, \alpha), (\alpha, \varnothing), (\varnothing, \alpha)\}$ (no discipline). Let $E(\pi^q_{nd,ns})$ denote the government payoff in this case (the subscript ns is used to emphasize that when $q > q^{\max}$ the public never supports implementation).

Fix $E \geq 0$. Consider any $p \in (0, 1)$, and let $L = (1 - p)E$, from (3). Expression (A6) shows that $E(\pi^G_{nd})$ is linear in p and in L. Thus, $E(\pi^G_{nd})$ remains linear in p after substituting $L = (1 - p)E$. More generally, the government payoff, seen as a function of p, is clearly always linear. In particular, $E(\pi^q_{nd,ns})$ is linear in p. We will show that $E(\pi^q_{nd,ns}) < E(\pi^G_{nd})$ when evaluated at both $p = 1$ and $p = 0$. Linearity then implies $E(\pi^q_{nd,ns}) < E(\pi^G_{nd})$ for all $p \in (0, 1)$. Since $E = L/(1 - p)$ was fixed at an arbitrary positive value, this implies $E(\pi^q_{nd,ns}) < E(\pi^G_{nd})$ for all $p \in (0, 1)$ and $L \geq 0$

Suppose that $p = 1$. In this case, protest blocks implementation with probability one, so the status quo always remains when $q > q^{\max}$. At the same time, the explicit loss caused by protest is zero, $L = 0$. It follows that $E(\pi^q_{nd,ns})$ takes on exactly the same value as in Case (i), where the government never selects implementation. Thus, $E(\pi^q_{nd,ns}) < E(\pi^G_d)$. Expressions (A4) and (A5) show that $E(\pi^G_d) = E(\pi^G_{nd})$ when evaluated at $p = 1$ and $L = 0$. We conclude that $E(\pi^q_{nd,ns}) < E(\pi^G_{nd})$.

Suppose that $p = 0$. Protest then never blocks implementation, and the explicit loss incurred by the government is $L = E$. Because the government selects the same policies as in the no discipline case, $E(\pi^q_{nd,ns})$ is equal to $E(\pi^q_{nd})$ evaluated at $p = 0$ and $L = E$, except that now the public protests implementation when the signals are both positive. Thus

$$E(\pi^q_{nd,ns}) = E(\pi^q_{nd}) - \left(\frac{1}{2}V + qR + \frac{1}{2}q^2V\right)E,$$

since $\Pr((\alpha, \alpha)^q) = \frac{1}{2}V + qR + \frac{1}{2}q^2V$. Using (A8), simple algebra yields

$$E(\pi^q_{nd,ns}) - E(\pi^G_{nd}) = -\frac{1}{2}VE - \frac{1}{2}q^2V\left[C_a - B + E - \sigma_-\right] - (1 - q)qV\left[C_a - B + E - \sigma_-\right] < 0,$$

where $C_a - B + E - \sigma_- > 0$ by Assumption 2. It follows that $E(\pi^q_{nd,ns}) < E(\pi^G_{nd})$.

Case (iii): the government selects implementation if and only if $\mathbf{s}^q = (\alpha, \alpha)$ (discipline). Let $E(\pi^q_{d,ns})$ denote the government payoff in this case.

Suppose $p = 1$. As in Case (ii), the status quo always remains when $q > q^{\max}$. The same logic as above then implies $E(\pi^q_{d,ns}) < E(\pi^G_d) = E(\pi^G_d)$.

Now suppose $p = 0$, so that protest never blocks implementation, and $L = E$. $E(\pi_{d,ns}^q)$ is equal to $E(\pi_d^q)$ except that now the public protests implementation when the signals are both positive. Thus

$$E(\pi_{d,ns}^q) = E(\pi_d^q) - \left(\frac{1}{2}V + qR + \frac{1}{2}q^2V\right)E.$$

Using (A4), and (A7), simple algebra yields

$$E(\pi_{d,ns}^q) - E(\pi_{nd}^G) = -\frac{1}{2}VE - \frac{1}{2}q^2V\left[C_a - B + E - \sigma_-\right] + (1-q)R\left[C_a - B + E - \frac{1}{2}\right],$$

and

$$E(\pi_{d,ns}^q) - E(\pi_d^G) = -\frac{1}{2}VE - \frac{1}{2}q^2V\left[C_a - B + E - \sigma_-\right] - qR\left[C_a - B + E - \frac{1}{2}\right].$$

Again, $C_a - B + E - \sigma_- > 0$ by Assumption 2. Thus, regardless of the sign of $\left[C_a - B + E - \frac{1}{2}\right]$, we have either $E(\pi_{d,ns}^q) < E(\pi_{nd}^G)$, or $E(\pi_{d,ns}^q) < E(\pi_d^G)$, or both. Linearity of the government payoff with respect to p then implies that either $E(\pi_{d,ns}^q) < E(\pi_{nd}^G)$, or $E(\pi_{d,ns}^q) < E(\pi_d^G)$, or both, for all $p \in (0,1)$. ∎

To prove Lemma 2 and Proposition 1, the following lemma is useful.

Lemma B1. *Suppose the government and the public share the same beliefs about the true state: $\sigma_{Gov} = \sigma_P$. Then, if the public will accept implementation, and $c/p < B$, then the government will select implementation. If the public will protest implementation, then the government will select implementation when $\sigma_{Gov} \geq C_a - B + E$.*

Proof of Lemma B1. Let $\sigma_{Gov} = \sigma_P = \sigma$. Recall that the public accepts implementation if $\sigma \geq C_a - c/p$. Assuming that the public accepts, the government selects implementation if $\sigma \geq C_a - B$. This latter condition is implied by $\sigma \geq C_a - c/p$ whenever $c/p < B$. Thus, if the public supports implementation, then the government also selects implementation. The second part of the lemma follows immediately from (2) and (3). ∎

Proof of Lemma 2. From Assumption 1 and the fact that $q \leq q^{\max}$, it follows immediately that the public accepts implementation if and only if both signals are positive. Assumption 2 implies that the government selects the status quo when both signals are negative. Because the public accepts implementation when both signals are positive, it also follows from Lemma B1 that the government must select implementation in that case if $c/p \leq B$.

It remains to show that the government also selects implementation when both signals are positive if $c/p \leq B$. Suppose the status quo was optimal. Then it must be that $\sigma_+^q < C_a - B$. This would imply $\sigma^q < C_a - B$, so the government would always choose the status quo. But the government could then earn strictly higher profits $E(\pi_d^G)$ by setting $q = 0$ and selecting implementation after two positive signals s shown in the proof of Lemma 1, which is a contradiction. ∎

Proof of Proposition 1. From Lemma 2, the government selects implementation in equilibrium when both signals are positive and the status quo when both signals are negative. Moreover, the public accepts implementation if and only if both signals are positive. It only remains to be shown is whether the government selects implementation or the status quo when the evidence is mixed.

Fix the value of $p \in (0,1)$, and let $E = L/(1-p)$ from (3). Suppose that $B < C_a - \frac{1}{2} + E$ (cases (i) and (ii)). Recall that $\sigma^q = \Pr(A|(\alpha, \varnothing)^q) \leq \frac{1}{2}$. Because $\sigma^q \leq \frac{1}{2} < C_a - B + E$, Lemma B1 implies that for all q's the government selects the status quo when the evidence is mixed. Thus, if $B < C_a - \frac{1}{2} + E$, for all q's the government is disciplined by public opinion.

Next, we derive the optimal q when $B < C_a - \frac{1}{2} + E$ (cases (i) and (ii)). Recall that $E(\pi_d^q)$ denotes the government payoff when the government is disciplined by public opinion and the size of the bias is q. Define

$$\Delta \pi_{d,nd}^q \equiv E(\pi_d^q) - E(\pi_{nd}^G), \quad \text{where } q \in [0, q^{\max}],$$

The optimal q solves

$$\max_{q \in [0, q^{\max}]} E(\pi_d^q),$$

or equivalently

$$\max_{q \in [0, q^{\max}]} \Delta \pi_{d,nd}^q,$$

since $E(\pi_{nd}^G)$ is independent of q. Thus the equilibrium level of bias is $q^* = \arg\max_{q \in [0, q^{\max}]} E(\pi_d^q)$.

From (A7) in Appendix A, we have

$$\Delta \pi_{d,nd}^q = RL - \frac{1}{2}q^2 V [C_a - B - \sigma_-] + (1 - p - q) R \left[C_a - \frac{1}{2} - B\right].$$

Assumption 2 implies $C_a - B - \sigma_- > 0$. Thus simple algebra yields

$$q^* = \begin{cases} \frac{R}{V} \frac{1/2 - C_a + B}{C_a - B - \sigma_-} & \text{if } B \in (C_a - \frac{1}{2}, C_a - \frac{1}{2} + E) \\ 0 & \text{if } B \leq C_a - \frac{1}{2} \end{cases}$$

36

The government indeed finds it optimal select implementation after two positive signals when $q = q^*$, so $\sigma^{q^*} > C_a - B$. If that were not the case, then the government would never select implementation. By Assumption 2, it could then earn strictly higher profits $E(\pi_d^G)$ by setting $q = 0$ and implementing after two positive signals, contradicting the fact that $q^* > 0$ was optimal.

Note that given condition (5), we restrict attention to parameter values for which the constraint $q \leq q^{\max}$ does not bind. Using the above formula for q^*, and the fact that $C_a - c/p = \sigma_+^{q^{\max}}$, we can rewrite condition (5) more explicitly as

$$\underbrace{\frac{R}{V} \frac{1/2 - C_a + B}{C_a - B - \sigma_-}}_{q^*} < \underbrace{\frac{-\frac{R}{V}\left(C_a - \frac{c}{p} - 1/2\right) + \sqrt{\left(\frac{R}{V}\right)^2 \left(C_a - \frac{c}{p} - 1/2\right)^2 - \left(C_a - \frac{c}{p} - \sigma_-\right)\left(C_{a-\frac{c}{p}} - \sigma_+^G\right)}}{\left(C_a - \frac{c}{p} - \sigma_-\right)}}_{q^{\max}}.$$

(B1)

Define $\overline{B}_1 \in (C_a - \frac{1}{2}, C_a - \sigma_-)$ such that (B1) holds strictly if and only if $B < \overline{B}_1$. Hence there exist values of $B \in (C_a - \frac{1}{2}, C_a - \frac{1}{2} + E)$ that satisfy condition (5). (If instead $B \leq C_a - \frac{1}{2}$, then $q^* = 0$ and assumption (5) always holds.) This proves parts (i) and (ii) of Proposition 1.

To prove part (iii), let $B \geq C_a - \frac{1}{2} + E$. Two cases can arise, depending on whether q is above or below the threshold $\hat{q}$, defined as the solution to $\sigma^q = C_a - B + E$. Recall that $\sigma^q = \frac{\frac{R}{2} + q(1-\theta)^2}{R + qV}$. Then

$$\hat{q} = \frac{R\left(\frac{1}{2} - C_a + B - E\right)}{V\left(C_a - B + E\right) - (1-\theta)^2} = \frac{R}{V} \frac{1/2 - C_a + B - E}{C_a - B + E - \sigma_-}. \tag{B2}$$

The requirement that $B \geq C_a - \frac{1}{2} + E$ implies $\hat{q} \geq 0$, and $E > 0$ implies $\hat{q} < q^* < q^{\max}$.

When $q < \hat{q}$, it is optimal for the government to select implementation when the signals are mixed, because $q < \hat{q}$ implies $\sigma^q > C_a - B + E$ (see Lemma B1). Conversely, when $q > \hat{q}$, it is optimal for the government to selects the status quo. Thus, when $B \geq C_a - \frac{1}{2} + E$, the normalized government payoff is

$$\Delta\pi_{LB}^q \equiv \begin{cases} \Delta\pi_{nd,nd}^q = E(\pi_{nd}^q) - E(\pi_{nd}^G) & \text{if } q \in [0, \hat{q}] \\ \Delta\pi_{d,nd}^q = E(\pi_d^q) - E(\pi_{nd}^G) & \text{if } q \in (\hat{q}, q^{\max}] \end{cases},$$

The optimal q solves

$$\max_{q \in [0, q^{\max}]} \Delta\pi_{LB}^q.$$

37

By the definition of $\hat{q}$, we have $\Delta\pi^q_{nd,nd} = \Delta\pi^q_{d,nd}$ when evaluated at $q = \hat{q}$. Looking at (A7) and (A8) then confirms that $\Delta\pi^q_{LB}$ is continuous on $[0, q^{\max}]$, since $\Delta\pi^q_{nd,nd}$ and $\Delta\pi^q_{d,nd}$ are both continuous in q on their respective domains.

From (A7), $\Delta\pi^q_{d,nd}$ is strictly concave in q, with

$$\frac{\partial^2 \Delta\pi^q_{d,nd}}{\partial q^2} = -V(C_a - B - \sigma_-) < 0,$$

and achieves its maximum at $q^* = \frac{R}{V}\frac{1/2-(C_a-B)}{C_a-B-\sigma_-} > \hat{q}$. It follows that the government maximizes $\Delta\pi^q_{LB}$ either by choosing discipline and $q = q^*$, or by choosing no discipline and some $q^L \leq \hat{q} < q^*$. The statement of Proposition 1 uses the notation $q^L(p, E)$ to emphasize the fact that the value of q^L may depend on both p and L, or equivalently on p and E.

We now show that given $L = (1 - p)E$, we have $q^L = 0$ whenever $p \leq 1/2$. Substituting $L = (1 - p)E$ into (A8) and differentiating twice with respect to q yields

$$\frac{\partial^2 \Delta\pi^q_{nd,nd}}{\partial q^2} = -V(C_a - B - \sigma_-) + 2V\left[(1-p)(C_a - B - \sigma_- + E)\right].$$

It follows that $\Delta\pi^q_{d,nd}$ is strictly convex if

$$p < 1 - \frac{1}{2}\left(\frac{C_a - B - \sigma_-}{C_a - B - \sigma_- + E}\right). \tag{7}$$

Thus, when $p \leq 1/2$, $\arg\max_{q\in[0,\hat{q}]} \Delta\pi^q_{nd,nd}$ is either 0 or $\hat{q}$. However, $\hat{q}$ yields a lower payoff than q^* since $\Delta\pi^{\hat{q}}_{nd,nd} = \Delta\pi^{\hat{q}}_{d,nd} < \Delta\pi^{q^*}_{d,nd}$, by $q^* > \hat{q}$. Thus, on $[0, q^{\max}]$, the optimal q is either 0 or q^*, depending on whether $E(\pi^0_{nd}) \lessgtr E(\pi^{q^*}_d)$ (or, equivalently, $\Delta\pi^0_{nd,nd} \lessgtr \Delta\pi^{q^*}_{d,nd}$). Finally, since $\overline{B}_1 > C_a - 1/2$, there exist values of $B > C_a - 1/2 + E$ that satisfy condition (5), provided that E is sufficiently small. $\blacksquare$

Proof of Proposition 2. Suppose first that $B \leq C_a - \frac{1}{2}$. Then by Proposition 1(i), the equilibrium level of bias is zero. Imposing $q = 0$ therefore leaves the public's payoff unchanged.

Suppose instead that $B \in (C_a - \frac{1}{2}, C_a - \frac{1}{2} + E)$. Then by Proposition 1(ii), the government is disciplined by public opinion and the equilibrium level of bias is $q^* = \frac{R}{V}\frac{1/2-C_a+B}{C_a-B-\sigma_-}$. Substituting $L = (1 - p)E$ into (A7) and imposing $q = 0$, the government will remain disciplined by public opinion since

$$\Delta\pi^G_{d,nd} = E(\pi^G_d) - E(\pi^G_{nd}) = R(1-p)\left(C_a - B + E - \frac{1}{2}\right) > 0.$$

Let $E(U_d^q)$ denote the public's payoff under discipline with bias q, given by (A10). We have

$$\Delta U_{d,d}^q \equiv E(U_d^q) - E(U_d^G) = -qR\left(C_a - \frac{1}{2}\right) - \frac{1}{2}q^2V\left(C_a - \sigma_-\right),$$

which is decreasing in q. Hence imposing $q = 0$ when $B \in (C_a - \frac{1}{2}, C_a - \frac{1}{2} + E)$ will strictly increase the public's payoff.

Now suppose that $B \geq C_a - \frac{1}{2} + E$. Proposition 1(iii) then implies that for $E(\pi_d^{q^*}) \leq E(\pi_{nd}^{q^L(p,E)})$, the government chooses no discipline, and the equilibrium level of bias is $q^L(p, E)$. Since $0 \leq q^L(p, E) \leq \hat{q}$, the government continues to choose no discipline when $q = 0$.

Let $\Delta U_{nd,nd}^q = E(U_{nd}^q) - E(U_{nd}^G)$ denote the normalized public payoff under no discipline, given by (A12). Differentiating (A12) with respect to q and rearranging yields

$$\frac{\partial \Delta U_{nd,nd}^q}{\partial q} = -R\left[p(C_a - \frac{1}{2}) - c\right] - 2qV\left[p(C_a - \sigma_-) - c\right] - V\left[(1-p)(C_a - \sigma_-) + c\right],$$

which is strictly negative by Assumption 1 and $\sigma_- < \sigma^G = 1/2$. It follows that when $E(\pi_d^{q^*}) \leq E(\pi_{nd}^{q^L(p,E)})$, imposing $q = 0$ will strictly increase the public's payoff if $q^L(p, E) > 0$, and leave the public's payoff unchanged if $q^L(p, E) = 0$.

If instead $E(\pi_d^{q^*}) > E(\pi_{nd}^{q^L(p,E)})$, then the government is disciplined by public opinion and the equilibrium level of bias is q^*. Moreover, substituting $L = (1-p)E$ into (A7), $B \geq C_a - \frac{1}{2} + E$ implies $\Delta \pi_{d,nd}^G \leq 0$. Hence imposing $q = 0$ leaves the government undisciplined by public opinion. By (A13), this will strictly decrease the public's payoff if

$$\Delta U_{d,nd}^{q^*} = Rc - \frac{1}{2}q^{*2}V\left(C_a - \sigma_-\right) + (1 - q^* - p)R\left(C_a - \frac{1}{2}\right) > 0.$$

Direct substitution yields $\Delta U_{d,nd}^G > 0$ and $\Delta U_{d,nd}^1 < 0$, where $\Delta U_{d,nd}^{q^*}$ is continuous and decreasing in q^*. Moreover, $q^* = \frac{R}{V}\frac{1/2 - C_a + B}{C_a - B - \sigma_-}$ is continuous and increasing in B, with $q^* = 0$ when evaluated at $B = C_a - \frac{1}{2}$. Define $\overline{B}_2 > C_a - \frac{1}{2}$ as the value of B for which $\Delta U_{d,nd}^{q^*} = 0$, where $\Delta U_{d,nd}^{q^*} > 0$ if and only if $B < \overline{B}_2$. Moreover, define $\overline{B} = \min(\overline{B}_1, \overline{B}_2)$, where (B1) holds strictly if and only if $B < \overline{B}_1$. It follows that, over the parameter region for which condition (5) does not bind, imposing $q = 0$ will strictly decrease the public's payoff if and only if both $B \in [C_a - \frac{1}{2} + E, \overline{B})$ and $E(\pi_{d,nd}^{q^*}) > E(\pi_{d,nd}^{q^L(p,E)})$.

First suppose that p is sufficiently small so that (7) holds (in particular this is the case for all $p \in (0, 1/2]$. Then Proposition 1 implies $q^L(p, E) = 0$, so that $E(\pi_d^{q^*}) > E(\pi_{nd}^{q^L(p,E)})$ is

equivalent to

$$\Delta\pi_{d,nd}^{q^*} = R(1-p)E - \frac{1}{2}q^{*2}V\left[C_a - B - \sigma_-\right] + (1 - q^* - p)R\left[C_a - \frac{1}{2} - B\right] > 0,$$

using (A7) and $L = (1-p)E$. Fix $B > C_a - \frac{1}{2}$. Note that $q^* = \frac{R}{V}\frac{1/2 - C_a + B}{C_a - B - \sigma_-}$ is independent of E, $\Delta\pi_{d,nd}^{q^*}$ is continuous and strictly increasing in E for all $p < 1$, and $\Delta\pi_{d,nd}^{q^*} < 0$ when evaluated at $E = 0$. Define $\underline{E}(B,p)$ as the value of E for which $\Delta\pi_{d,nd}^{q^*} = 0$:

$$\underline{E}(B,p) = \frac{\frac{1}{2}q^{*2}V\left[C_a - B - \sigma_-\right] + (1 - q^* - p)R\left[B - C_a + \frac{1}{2}\right]}{R(1-p)}, \tag{8}$$

so that $\Delta\pi_{d,nd}^{q^*} \le 0$ for all $E \le \underline{E}(B)$, and $\Delta\pi_{d,nd}^{q^*} > 0$ for all $(\underline{E}(B), B - C_a + \frac{1}{2}]$. To show that $\underline{E}(B,p) < B - C_a + \frac{1}{2}$, notice that $\underline{E}(B)$ is decreasing in q^* over the interval $[0, \frac{R}{V}\frac{1/2 - C_a + B}{C_a - B - \sigma_-}]$, and $\underline{E}(B,p) = B - C_a + \frac{1}{2}$ when evaluated at $q^* = 0$. Differentiating (8) with respect to p simplifying shows that

$$\frac{\partial \underline{E}(B,p)}{\partial p} = \frac{-q^*}{2R(1-p)^2}\left(\frac{1}{2} - C_a - B\right) < 0.$$

Now suppose that $p \ge 1/2$ is sufficiently large to violate (7). We showed above that $E(\pi_{nd}^G) > E(\pi_d^{q^*})$ when $E = 0$. The optimality of $q^L(p, E)$ implies $E(\pi_{nd}^{q^L(p,E=0)}) \ge E(\pi_{nd}^G)$, so it follows that $E(\pi_{nd}^{q^L(p,E=0)}) > E(\pi_d^{q^*})$.

(A4) shows that $E(\pi_d^q)$ is independent of p and E. Using $L = (1-p)E$, (A5) shows that $\frac{\partial}{\partial E}E(\pi_{nd}^q) < 0$ for all $q < 1$, and also that $B > C_a - \frac{1}{2} + E$ implies $\frac{\partial}{\partial p}E(\pi_{nd}^q) < 0$. Moreover, $\frac{d}{dE}E(\pi_{nd}^{q^L(p,E)}) = \frac{\partial}{\partial E}E(\pi_{nd}^{q^L(p,E)})$, and $\frac{d}{dp}E(\pi_{nd}^{q^L(p,E)}) = \frac{\partial}{\partial p}E(\pi_{nd}^{q^l(p,E)})$, by the envelope theorem and the optimality of $q^L(p, E)$, since $\frac{\partial}{\partial q}E(\pi_{nd}^q) = 0$ when evaluated at $q^L(p, E)$. It follows that $E(\pi_{nd}^{q^L(p,E)})$ is decreasing in E and p.

Define $\underline{E}(B,p)$ as the value of E for which $E(\pi_{nd}^{q^L(p,E)}) = E(\pi_d^{q^*})$. We showed above that there exists such a $\underline{E}(B,p)$ for any p sufficiently small to satisfy (7), with $\underline{E}(B,p) < B - C_a + 1/2$. Thus, by $\frac{d}{dE}E(\pi_d^{q^L(p,E)})$ and $\frac{d}{dp}E(\pi_d^{q^L(p,E)})$, there must also exist such a $\underline{E}(B,p) < B - C_a + 1/2$ for any p sufficiently large to violate (7), with $\frac{\partial}{\partial p}\underline{E}(B,p) < 0$.

Now fix $E \in (0, B - C_a + \frac{1}{2})$, and let p tend to 1. From (A8), taking the first order condition for $E(\pi_{nd}^q)$ with respect to q yields the optimal bias under no discipline

$$q^L(p,E) = \frac{R\left[p(\frac{1}{2} - C_a + B) + (1-p)E\right] - V\left[(1-p)(C_a - B - \sigma_- + E)\right]}{V\left[(C_a - B - \sigma_-) - 2(1-p)(C_a - B - \sigma_- + E)\right]}, \tag{9}$$

so that in the limit $q^L(p, E)$ tends to q^*. The proof of Proposition 1 showed that $q^* > \hat{q}$, given by (B2), for any $E > 0$. Hence, by the definition of $\hat{q}$, $q^* > \hat{q}$ implies $E(\pi_{nd}^{q^*}) < E(\pi_d^{q^*})$,

which in turn yields $E(\pi_{nd}^{q^L(p,E)}) < E(\pi_d^{q^*})$. This is the case for any fixed $E \in (0,1)$, so it follows that $\lim_{p \to 1} \underline{E}(B,p) = 0$. ∎

Proof of Proposition 3. Standard results in decision theory imply that, for any *given* belief about q that the public may hold, setting $q = 0$ is a weakly dominant strategy for the government (see Marschak and Radner, 1972, pp. 65-67. Further details are available from the authors upon request). It follows that the equilibrium bias is $q = 0$. Notice also that when both signals are negative, the government selects the status quo since $C_a - B > \sigma_-$ by Assumption 2.

For case (i), the government has a dominant strategy to select the status quo when the evidence is mixed. If the public does not protest, then the government earns $-C_a/2 + B$ from implementation and $-C_n/2$ from the status quo, where the status quo is optimal by $B \leq C_a - 1/2$. This in turn implies $C_a - B - 1/2 + E \geq 0$, so the status quo is also optimal if the public were to protest implementation. The public therefore believes $\sigma_p = \sigma_+^G$ upon observing implementation. It implementation because $C_a - c/p \leq \sigma_+^G$ by Assumption 1.

For cases (iv) and (v), the government has a dominant strategy to select implementation when the evidence is mixed. Say the evidence is mixed. If the public protests, then the government earns $(1-p)(-C_a/2 + B) + p(-C_n/2) - L$ from implementation and $-C_n/2$ from the status quo, where implementation is optimal by $B \geq C_a - 1/2 + E$. This in turn implies $B \geq C_a - 1/2$, so implementation is also optimal if there public were to accept it. The public therefore believes the probability of the good state conditional on observing implementation is

$$\frac{[1 - (1-\theta)^2]}{[1 - (1-\theta)^2] + 1 - \theta^2} = \frac{\theta^2 + R}{1 + R} \equiv \hat{\sigma}.$$

Given beliefs $\sigma_p = \hat{\sigma}$, the public earns $(1 - \hat{\sigma})(-C_a)$ from accepting implementation and $(1-p)(1 - \hat{\sigma})(-C_a) + p\hat{\sigma}(-C_n) - c$ from protesting. It follows that the public will accept implementation if and only if $\hat{\sigma} \geq C_a - c/p$.

For case (iii), if $\hat{\sigma} \geq C_a - c/p$, then the public will accept implementation. The government therefore earns $-C_a/2 + B$ from implementation and $-C_n/2$ from the status quo when the evidence is mixed. It follows that implementation is optimal because $B > C_a - 1/2$.

For case (ii), we have $B \in (C_a - \frac{1}{2}, C_a - \frac{1}{2} + E)$ and $\hat{\sigma} < C_a$. Suppose the government always selects implementation when the evidence is mixed. Then the public will protest implementation since $\hat{\sigma} < C_a$. But then the government could profitably deviate to the

status quo when the evidence is mixed, since $B < C_a - \frac{1}{2} + E$, in contrast to cases (iv) and (v).

Suppose instead the government always selects the status quo when the evidence is mixed. Then the public will accept implementation since $\sigma_+^G \geq C_a - c/p$ by Assumption 1. But then the government could profitably deviate to implementation when the evidence is mixed since $B > C_a - \frac{1}{2}$. It follows that any equilibrium must be in mixed strategies.

Let $\check{z}$ denote the probability that the public protests conditional on observing implementation, and let $\check{s}$ denote the probability that the government selects implementation when the evidence is mixed. The indifference condition for the government is

$$-\frac{C_n}{2} = (1 - \check{z})\left[\frac{-C_a}{2} + B\right] + \check{z}\left[(1 - p)(\frac{-C_a}{2} + B) + p(\frac{-C_n}{2}) - L\right],$$

$$\implies \quad \check{z} = \frac{\frac{1}{2} - C_a + B}{p(\frac{1}{2} - C_a + B) + (1 - p)E}.$$

The indifference condition for the public is $\Pr(A \mid a, \check{s}) = C_a - c/p$, where $\Pr(A \mid a, \check{s})$ is the public's belief that the state is good, $S = A$, when the government selects implementation $(p = a)$, given that the government implements with probability one if $(\alpha, \alpha)^G$, with probability $\check{s}$ if the evidence is mixed, and with probability zero if $(\varnothing, \varnothing)^G$. Using Bayes' rule

$$\Pr(A \mid a, \check{s}) = \frac{\theta^2 + \check{s}R}{V + 2\check{s}R} = C_a - \frac{c}{p} \implies \check{s} = \frac{\theta^2 - (C_a - \frac{c}{p})V}{R\left(2(C_a - \frac{c}{p}) - 1\right)} = \left(\frac{V}{2R}\right)\frac{\sigma_+^G - C_a + \frac{c}{p}}{C_a - \frac{c}{p} - 1/2}.$$

It is easy to show that $\hat{\sigma} < C_a - c/p$ implies $\check{s} < 1$ and that $B \in (C_a - \frac{1}{2}, C_a - \frac{1}{2} + E)$ implies $\check{z} < 1$.

To complete the proof, notice that the government selects implementation with positive probability for all cases (i)-(v), when the evidence is mixed. Thus, implementation is always a best response when the evidence is mixed, given the strategy of public. It follows that implementation is also a best response when both signals are positive, by $\sigma_+^G > \sigma^G = 1/2$, so that in equilibrium the government will select implementation after a positive report. ∎

Proof of Proposition 4. Obvious. ∎

Proof of Proposition 5. Note that Constitutions III and IV yield the same citizen welfare since nondisclosure implies $q = 0$ in equilibrium (Proposition 3). Moreover, by Proposition 4 Constitution I dominates Constitution III (and hence also Constitution IV). Thus disclosure

is always a feature of an optimal constitution. Finally, Propositions 1 and 2 demonstrate that either Constitution I or II can be optimal. ∎