Abstract

Recent papers in political accountability have documented the detrimental behavior that results when voters observe an incumbent’s actions or their outcomes because the incumbent distorts her actions in order to signal competence; however, signaling requires private information. We consider a model in which all information is public, and voters observe actions and outcomes perfectly. We call this situation full transparency. Incumbents are either competent or incompetent. Compared to the status quo, voters are made better off when competent types implement policy, but worse off when incompetent types do so. Under full transparency, only those who are sufficiently unlikely to be competent, according to the public signal, choose to implement their ideal policy. Those who are likely to be competent prefer the status quo. In contrast, private information provides incentives for those more likely to be competent to implement policy, and some of those least likely to be competent to choose the status quo. Consequently, private information may provide higher welfare than full transparency.
1 Introduction

Consider the dilemma faced by an incumbent mayor in a city that is considering raising the minimum wage to fifteen dollars per hour. Neither the voters nor the mayor know for sure which the best course of action is for the local economy. The public believes that raising the minimum wage is not appropriate, based on whatever information is publicly available. However, the mayor has additional private information, such as reports from the city’s economic advisor, that the minimum wage should be raised. The mayor’s belief that the wage should be raised is more accurate on average than that of the voters, because of the mayor’s additional information. Hence, the best course of action for the welfare of the voters would be for the incumbent to raise the minimum wage. Unfortunately, if the mayor chooses to raise the minimum wage, then voters are likely to see the mayor as incompetent, because they believe raising the minimum wage is the wrong course of action. Hence, the mayor has an electoral incentive to ignore her private information in order to send a signal that she is competent.

But what if instead, the city’s advisor released the report publicly, so that there was no private information on the part of the incumbent? The mayor and the voters would be equally informed about the merits of raising the minimum wage, and there would be no possibility of signaling. It would seem that the incumbent should now be free to do whatever is best for the city.

This example motivates the following question: If all information is public, and voters observe the incumbent’s actions and their consequences, do incumbents implement policy in a welfare maximizing way?

Transparency is the ability of voters to observe the actions and consequences of politicians’ actions. As shown in our example, this may be welfare impairing. We refer to the situation in which actions, consequences, and signals are observed as full transparency. To
investigate the properties of full transparency; we consider a two period model of an incumbent with an opportunity in each period to implement one of her ideal policies. In the first period, both voters and incumbents get a signal of the probability that the policy will deliver good outcomes. The incumbent then chooses between implementing her ideal policy and the status quo to maximize her expected probability of reelection. The status quo reveals no further information about her competence, while implementation of her ideal policy reveals her competence with certainty. Voters observe the incumbent’s choice and its outcome, if applicable, and update their beliefs about the incumbent. The belief about the competence of the challenger is then drawn from a unimodal CDF, and voters reelect the incumbent if and only if the belief about her is higher than that of the challenger. We then consider the case when the signal is not revealed to the voters. If the incumbent chooses the status quo with private information, then voters must vote on the basis of the expected signal conditional on observing the status quo. This informational change implies that the probability of reelection after choosing the status quo is now common across types, rather than type specific.

The first main result is that incumbents fail to behave optimally in the full transparency equilibrium, and may even minimize welfare. Incumbents with high signals are risk averse and choose the status quo to avoid electoral risk, while incumbents with low signals are risk loving and gamble by implementing their ideal policy. In private information, welfare may be strictly higher.

The difficulty is that it is impossible to separate information about the quality of the policy ideas from the quality of the person who had them. Returning to our example, suppose that raising the minimum wage was the mayor’s idea to begin with. If it turns out that the policy is very likely to fail, then this reflects poorly on the competence of the mayor. If it turns out that the policy is very likely to work, then the mayor appears very likely to be competent. If the mayor appears incompetent, she is very likely to lose the election. Her
only hope is to pass the minimum wage increase despite the signals against it and hope to be proven right, so that voters will see her as competent. While the policy is not likely to work, it is a gamble she is willing to take. In contrast, when the mayor already appears competent because her minimum wage increase is likely to be good for the local economy, there is a risk that upon actually implementing the policy, it does not work and she looks incompetent. While it is very likely to work out, the gains in that event are minimal, while the losses from a failure of the policy are severe. In this case, the incumbent would prefer to avoid the gamble.

The second main result is that when the incumbent’s signal remains private, if the status quo is chosen, voters only have an expected signal to evaluate the incumbent with. For those with low signals, this may be more appealing than taking a policy gamble, while those with high signals feel they can do better by taking a policy gamble. In effect, incumbents with a low signal have an incentive to hide their bad signal by choosing the status quo, and incumbents with a high signal have an incentive to prove they received a high signal by implementing their ideal policies.

In our example, when information is private, voters will realize that if the mayor refuses to implement the wage increase, it must be that the report was not favorable. They infer from this that whenever the mayor chooses not to act on her idea, it must be that the idea was bad, and hence, that she is incompetent. When the mayor receives a strong signal in favor of the wage increase in this environment, she has an incentive to act on it, to avoid the lackluster chance of re-election associated with inaction. Consequently, voters get more policy changes from those who are likely to be making improvements, while they get less from those who are likely to make things worse.

In Section 2, we review the related literature. In Section 3, we detail the assumptions of the full transparency model. In Section 4, we characterize the full transparency equilibrium. In Section 5, we compared the full transparency equilibrium to the optimal behavior from the
perspective of voters. In Section 6, we detail private information scenarios and their welfare implications. In Section 7, we summarize the main results and discuss possible extensions.

2 Related Literature

This paper relates to the literature on transparency of action and consequence. Canes-Wrone, Herron, and Shotts (2001) show that an incumbent with an imperfect private signal may sometimes *pander* by choosing to do what voters think is best, despite having private information to the contrary, in order to appear competent. This kind of behavior is ascribed by Prat (2005) to the “wrong kind” of transparency, that is, the observation of actions rather than outcomes alone. When voters can observe actions, politicians have an incentive to choose actions that voters will perceive as competent, rather than simply seeking to generate good outcomes. However, even if only outcomes are observable, politicians may still make bad decisions to maximize the probability of a good outcome, rather than maximizing the expected outcome (Fox and Weelden 2015). It need not be the case that politicians distort their behavior toward what voters’ preferred policies either. Instead, politicians may *posture* by choosing bold actions that are very unlikely to be appropriate (Fox and Stephenson 2011).

In all of these models, detrimental incumbent behavior arises from her desire to signal. This paper adds to this literature by determining if incumbent behavior on risky policy decisions is improved without private information, and explores a novel way in which private information may actually enhance welfare. Further, transparency in the sense explored by the existing literature does not inform us on the effects of disclosing previously confidential information. For example: Do public reports by a government agency about the projected impact of a new policy make voters better off than if the reports were sent only to the policymaker?

Hortala-Vallve (2017). The authors consider incumbents who do not have private information about the incumbent’s competence on a reform task, and assume voters are able to set the reelection probability associated with the status quo. The incumbent strategically chooses to implement reform to increase her expected reelection probability, and consequently, may over- or under-invest in risky reform. This suggests that even without private information, inefficiency may arise. When a challenger can launch an informative campaign about the incumbent, welfare deteriorates further. We add to this analysis by considering reelection probabilities determined by the reputation of the incumbent and the distribution of challengers. Further, we analyze how private information, rather than more informative campaigns by the challenger, may improve welfare.

Similar ideas are explored in the literature related to Bayesian Persuasion. There, incumbents may experiment with policy in order to manipulate learning about the state of the world, which then makes the incumbent’s platform more or less desirable. When an incumbent can choose the degree of experimentation by controlling the extremism of the implemented policy, a negative valence shock can encourage more extensive policy changes (Izzo 2018). Similarly, when the incumbent can design an optimal experiment, the informativeness of this experiment is decreasing in the incumbent’s valence (Alonso and Câmara 2016). However, the outcomes of the experiments in both models are unrelated to the incumbent’s quality, and Alonso and Câmara (2016) use the statistical independence of valence from experimental outcomes extensively in their proofs. The present analysis complements this work by exploring which are directly dependent upon quality of the incumbent, which may imply a very different relationship between current beliefs about her and her risk attitudes. Further, we consider the case in which the incumbent has only a single experiment which they can implement or not, rather than the ability to create optimal experiments, and find similar results with this restriction.

This paper also relates to the larger literature on reputation-concerned decision makers,
as initiated by Holmström (1999). This paper considers a novel utility function for reputation, the CDF of other’s reputations. Decisionmakers’ future employment prospects depend not only on their absolute reputations, but on their reputations relative to other potential applicants. As such the decision maker’s utility from reputation may be convex for low values and concave for higher values. Hence, the decision maker’s risk attitudes are contingent upon their initial reputation and the specific lottery of possible outcomes. As we show, this implies risk loving low reputational types and risk averse high reputational types, both of which act contrarily to maximizing the expected outcome for the principal. While there has been some exploration of policy implementation under career concerns, such as Fu and Li (2014), these models consider private information. We show that distinct issues arise in the absence of private information, and that over and under investment in a risky project can exist simultaneously.

3 Model Setup

There are two periods, \( t \in \{1, 2\} \), with an election after period one. The players are an incumbent politician, a challenger, and a representative voter. Politicians are referred to by feminine pronouns and the representative voter by masculine pronouns. Each period, the incumbent has an opportunity to change policy on one issue. The incumbent politician in period \( t \) has a known ideal policy, \( \theta_t \), for the issue of opportunity in period \( t \), different from the status quo on that issue, \( s_t \). She may either implement her ideal policy \( (x_t = \theta_t) \), or leave the status quo \( (x_t = s_t) \). The representative voter, on the other hand, is uncertain about his preference for \( \theta_t \) relative to \( s_t \). If the incumbent is competent \( (\tau = H) \), then her ideal policy will make the voter better off than the status quo. If the incumbent is incompetent \( (\tau = L) \), then her ideal policy will make the voter worse off than the status quo.\(^2\)

\(^2\) Izzo (2018) uses a similar framework, in which the incumbent has a known bliss point and uses policy strategically to maximize the probability that the voter prefers it, but with a continuous choice rather than a
leaves the status quo policy in place, then the voter’s utility is unchanged. Formally, let the voter’s utility be defined by

\[
v_1(x_1, \tau) = \begin{cases} 
1 & \text{if } x_1 = \theta_1, \tau = H \\
0 & \text{if } x_1 = s_1 \\
-1 & \text{if } x_1 = \theta_1, \tau = L.
\end{cases}
\]

In the second period, voter utility is scaled so that \( v_2 = \lambda \) if a competent type implements her ideal policy, and \( v_2 = -\lambda \) if an incompetent type implements her ideal policy. The parameter \( \lambda \) measures the importance of the second period relative to the first period, which determines the value of learning about the incumbent’s type. Formally, second period voter utility is given by

\[
v_2(x_2, \tau, \lambda) = \begin{cases} 
\lambda & \text{if } x_2 = \theta_2, \tau = H \\
0 & \text{if } x_2 = s_2 \\
-\lambda & \text{if } x_2 = \theta_2, \tau = L.
\end{cases}
\]

Uncertainty about the outcomes of policy \( \theta \) is equivalent to uncertainty about the the competence of the incumbent. Before policy is chosen, both the incumbent and the voter receive a signal indicating the probability that the policy will improve the voter’s utility, or \( P(v_1(\theta_1) = 1) \). Because policy will succeed if and only if the incumbent is competent, it must be the case that \( P(v_1(\theta_1) = 1) = P(\tau = H) \). We refer to the belief that the incumbent is competent as her *reputation*, and denote its value before policy is chosen by \( b_0 \). The signal may be more or less informative in any particular realization, so that any \( b_0 \in [0, 1] \) may be received with positive probability. Specifically, \( b_0 \) is distributed according to a a continuous, discrete one. Harrington (1993) makes this assumption also, describing the incumbents as being “dogmatic.”
differentiable CDF $G$ with expectation $E(b_0) = \frac{1}{2}$, such that $g(b_0) > 0$, $\forall b_0 \in [0, 1]$. Because the incumbent’s choices are characterized by her initial reputation, we refer to $b_0$ as her reputation type.

If the incumbent implements her ideal policy, then it reveals her competence with certainty. If the incumbent chooses the status quo, then no further information is generated about her competence. The updated belief about the incumbent’s competence at the end of the first period is $b_1$. Formally, the reputation the incumbent at the time of the election is

$$b_1(v_1) = \begin{cases} 
1 & \text{if } V_1 = 1 \\
 b_0 & \text{if } V_1 = 0 \\
0 & \text{if } V_1 = -1.
\end{cases}$$

The challenger’s reputation is unknown when policy is chosen. At the end of period 1, before the election, a challenger with reputation $b_1^c$ is drawn from the continuous, differentiable CDF $F$, with support $[0, 1]$. We assume $F$ is strictly unimodal. That is, there exists a unique mode $\mu \in (0, 1)$, such that $F$ is strictly convex on $(0, \mu)$ and strictly concave on $(\mu, 1)$. We also assume the pdf satisfies $f(0) < 1$, $f(1) < 1$ and $f(b_1^c) > 0$, $\forall b_1^c \in (0, 1)$. The strict unimodality of $F$ also implies that $f$ is a strictly quasiconcave function. Denote the median of this distribution by $m$, and the mode by $\mu$.

Both the challenger and incumbent care about re-election and implementing ideal policies in a lexicographic way. Whenever one policy delivers a strictly higher probability of re-election, it is strictly preferred. If the two policies are equivalent in their probability of re-election, then the incumbent prefers $x_t = \theta_t$ to $x_t = s_t$. In the second period, the incumbent has no possibility of re-election regardless of her choice of $x_2$. Hence, whomever holds office in period two chooses $x_2 = \theta_2$, and thus the the second period expected utility

---

3. Alonso and Câmara (2016) make an identical assumption about the incumbent’s preferences.
given \( b_1 \) is

\[
\mathbb{E}(v_2|b_1) = b_1 \lambda + (1 - b_1)(-\lambda) = (2b_1 - 1)\lambda.
\]

The representative voter is forward looking, and re-elects the incumbent if and only if he expects better second period utility from the incumbent than the challenger. Therefore, the representative voter re-elects the incumbent if and only if \( b_1 \geq b_1^c \). The linearity of the voter’s expected second period utility in reputation also ensures that, if there is not full information about the incumbent’s signal, the voter can use the expected signal in its place. When we consider the case of private information, this will greatly simplify the analysis.

**Example.** Returning to our running example, the mayor of a city has an opportunity to change the minimum wage. Her ideal policy is to raise it to fifteen dollars. Currently, the city has only a nine dollar per hour minimum wage. The appropriateness of this policy depends on the incumbent’s understanding of the local economy. If the incumbent does not have a very strong grasp of the features of the local economy, then fifteen dollars per hour may be too high and lead to unemployment or lower growth. If the incumbent is well informed about the local economy, then this policy change is likely to raise incomes without much cost.

Voters do not know exactly how this change would influence prices, benefits, or employment. However, an independent research bureau within the city government releases an impact report, which may report that the minimum wage is certain to work, certain to fail, or anything in between. The report finds a 60% chance that this policy will be beneficial to the voter overall. The voter then infers that there is a 60% chance that the incumbent is a competent type. If the incumbent leaves the current minimum wage, then the voter learns nothing new about the effects of the higher minimum wage and hence, nothing new about the competence of the incumbent. When choosing whether to vote for the incumbent or the challenger, voters believe that, whatever the policy issue of opportunity may be in period two, the incumbent’s policy will benefit them with a 60% probability.
3.1 Timeline of play for the full transparency model

1. Nature chooses the competence of the incumbent, $\tau \in \{H, L\}$.

2. Nature sends the signal $b_0$ to both the voter and the incumbent.

3. The incumbent chooses to implement her ideal policy ($x_1 = \theta_1$) or the status quo ($x_1 = s_1$), and the voter observes this.

4. Voters and the incumbent observe the outcome of the incumbent’s choice, $v_1(x_1, \tau)$.

5. Voter updates her beliefs about the incumbent’s competence to $b_1(v_1)$.

6. Voters observe $b_1^c$ and choose to re-elect the incumbent or elect the challenger.

7. In the second period, the elected politician chooses to implement her ideal policy ($x_2 = \theta_2$).

4 Strategies Under Full Transparency

As described in the set up, the incumbent in the second period implements her ideal policy on the second period policy opportunity, and therefore voters elect the candidate with the higher reputation. Consequently, the incumbent would like to manipulate her reputation to improve her expected outcome in the election. Because policy outcomes are a function of her competence, the incumbent therefore has an incentive to strategically implement her ideal policy, in order to manipulate voter’s learning about her.

Because the incumbent does not know her competence, implementing her ideal policy is a gamble. If the policy is detrimental to the representative voter, the incumbent is shown to be incompetent, and therefore, is very unlikely to be re-elected. If the policy turns out to be beneficial to the representative voter, then the incumbent is shown to be competent, so
that they are very likely to be re-elected. Therefore, implementing policy inherently means accepting a lottery of possible electoral outcomes.

On the other hand, choosing the status quo ensures that the voters learn nothing new about the appropriateness of the incumbent’s ideal policy, so that her reputation remains unchanged. Voters cannot infer anything from her choice because the incumbent has no private information. Rather, the choice of the status quo is a form of signal jamming. It prevents new information from being created. While the incumbent has no ability to pander or posture, she is still in a position to control the flow of information about herself, by choosing to either implement her ideal policy or not.

The payoff to the incumbent of creating a given reputation is the probability of re-election associated with each reputation. Consequently, the value to the incumbent of a reputation of $b_1$ is $F(b_1)$. The CDF of challenger reputation in effect becomes the payoff function for the incumbent. When the incumbent chooses the status quo, $b_1$ is equal to $b_0$, and hence her payoff is $F(b_0)$. By choosing action, her payoff is the expected probability of re-election conditional on the initial reputation, $E(F(b_1)|b_0)$.

To get the intuition for when the incumbent prefers to implement her ideal policy, consider Figure 1. The fully informative nature of the policy outcome means that the lottery has only two outcomes, $(0, F(0))$ and $(1, F(1))$. Because the support of $F$ is $[0, 1]$, we know $F(1) = 1$ and $F(0) = 0$. Therefore, $E(F(b_1)|b_0)$ is represented by a line connecting $(0, 0)$ and $(1, 1)$. The incumbent only prefers to implement her ideal policy if $E(F(b_1)|b_0) > F(b_0)$, and this is only true when $b_0$ is sufficiently small; namely, to the left of the intersection point of the line and the CDF. The incumbent prefers the status quo if $F(b_0) > E(F(b_1)|b_0)$. This is true to the right of the intersection point. Hence, the incumbent only implements her ideal policy when $b_0$ is sufficiently low; otherwise, she is effectively risk averse and prefers to avoid taking a gamble.

To develop this formally, define $D(b_0) \equiv E(F(b_1)|b_0) - F(b_0)$, the difference in expected
Figure 1: An incumbent with reputation $b_0$ would prefer to implement her ideal policy.

re-election probability when $x_1 = \theta_1$ relative to when $x_1 = s_1$. The incumbent prefers $x_1 = \theta_1$ if and only if $D(b_0) \geq 0$ and prefers $x_1 = s_1$ if and only if $D(b_0) < 0$. Note that at $D(b_0) = 0$, the both alternatives give the incumbent equal re-election probabilities, and so she prefers to implement her ideal policy.

Let $p(b_0)$ be the probability that implementing her ideal policy results in a good outcome for the voter, given incumbent’s reputation, $b_0$. The expected utility for the incumbent of implementing policy, given $b_0$, is:

$$E(F\mid b_0) = p_+(b_0)F(1) + (1 - p_+(b_0))F(0)$$

Recall that $F(1) = 1$ and $F(0) = 0$. Therefore, $E(F\mid b_0) = p_+(b_0)$. Because policy succeeds if and only if the incumbent is competent, $p_+(b_0) = b_0$. Hence,

$$D(b_0) = b_0 - F(b_0).$$

Therefore, the incumbent prefers to implement her ideal policy if and only if $b_0 > F(b_0)$.

To determine when policy implementation is preferred to the status quo, note first that
\(D(0) = 0 \times (1) - F(0) = 0\) and \(D(1) = 1 \times (1) - F(1) = 0\). Differentiating \(D\) with respect to \(b_0\), we find that
\[
D'(b_0) = 1 - f(b_0).
\]

Hence, it must be the case that \(f'(b_0) = 1\) for any stationary point of \(D(b_0)\).

In Lemma 1, we prove that given a unimodal CDF, there is at most one more point where \(D(b_0) = 0\) in \((0, 1)\).

**Lemma 1.** There is at most one point \(b_{FT} \in (0, 1)\) such that \(D(b_{FT}) = 0\).

**Proof.** It has already been established that \(D(0) = 0\) and \(D(1) = 0\). Suppose there exist at least two additional roots in the interval \((0, 1)\), \(b^1\) and \(b^2\), chosen arbitrarily if there are more than two, labeled such that \(b^1 < b^2\). By Rolle’s Theorem, there must exist a stationary point of \(D(b_0)\) in each interval \((0, b^1), (b^1, b^2),\) and \((b^2, 1)\). Hence, there must exist at least three stationary points of \(D(b_0)\), each of which must satisfy \(f(b_0) = 1\). This is impossible by the strict quasi-concavity of \(f\), and therefore there is at most one stationary point. 

![Figure 2](image.png)

Figure 2: Unimodality implies there are at most two points where \(D'(b) = 0\).

Under the assumption that \(1 > f(1)\) and \(1 > f(0)\), there is a unique \(b_{FT}\) such that it is only preferred to implement her ideal policy when \(b_0 < b_{FT}\), as shown in Figure 3.

---

4. The cases in which this assumption fails are cases in which all incumbents choose the same action regardless of their reputation, either all choosing their ideal policies or the status quo.
Figure 3: An incumbent with any reputation $b_0 \leq b_{FT}$ would prefer to implement her ideal policy.

**Proposition 1.** In the full transparency model, if $1 > f(1)$ and $1 > f(0)$, there exists a unique $b_{FT} \in (0, 1)$ such that the incumbent strictly prefers $x_1 = \theta_1$ if $b_0 \leq b_{FT}$, and the incumbent strictly prefers $x_1 = s_1$ if $b_0 > b_{FT}$.

**Proof.** Because $1 > f(1)$ and $1 > f(0)$, it follows that $D'(0) > 0$ and $D'(1) > 0$. Recall that $D(0) = 0$ and $D(1) = 0$. Therefore, for a positive $b_0$ located arbitrarily close to 0, $D(b_0) > 0$. For $b_0 < 1$ but arbitrarily close to 1, $D(b_0) < 0$. Thus, by the Intermediate Value Theorem, there must exist at least one point $b_{FT} \in (0, 1)$ such that $D(b_{FT}) = 0$. By Lemma 1, there is at most one $b_{FT}$, and hence, there must be exactly one.

Further, because $D(b_0) < 0$ for $b_0$ arbitrarily close to 1 and $b_{FT}$ is the unique root of $D$ on $(0, 1)$, it must be the case that $D(b_0) < 0$ for every $b_0 > b_{FT}$, and by a symmetric argument, $D(b_0) > 0$ for every $b_0 < b_{FT}$.

Because the incumbent prefers to implement her ideal policy if and only if $D(b_0) \geq 0$ and the status quo if and only if $D(b_0) \leq 0$, the incumbent has a strict preference for $x_1 = \theta_1$ when $b_0 < b_{FT}$ and a strict preference for $x_1 = s_1$ when $b_0 > b_{FT}$.

Proposition 1 implies that when the signal about the incumbent’s policy is public knowl-
edge, only those with weak evidence that their actions will be successful choose to implement her ideal policy. While the incumbent cannot misrepresent her signal by the choice of policy, she can control the flow of new information by choosing to implement her ideal policy or not. New information is lucrative to the incumbent when she has a low chance of winning, and if voters are rational and trying to select good types, a low chance of winning comes from a signal that she is inept. Conversely, a candidate with a strong signal is already at an electoral advantage, and so she has no incentive to take a risk by implementing her ideal policy.

A fully informative lottery is a special case, and one which makes the gamble very stark for the incumbent. In reality, good policies may fail sometimes, and bad policies may occasionally appear to work. Voters may find the challenger to be unpalatable despite her talents, or like her despite her incompetence. Further, the incumbent may know the identity of the challenger, so that there is limited uncertainty about the challenger’s quality. In the appendix, we show formally that while the particular value of $b_{FT}$ may change, it remains the cases that only those with a sufficiently low $b_0$ prefer to implement their ideal policies, and those with a higher $b_0$ prefer the status quo.

The equilibrium strategy profile is summarized in Theorem 1.

**Theorem 1.** In the Perfect Bayesian Equilibrium, the incumbent chooses $x_1 = \theta_1$ if and only if $b_0 \leq b_{FT}$, and $x_t = s$ otherwise. The representative voter re-elects the incumbent if and only if $b_1(y_1) \geq b_1^c$.

In the special case in which $F$ is also symmetric, then because the density is strictly positive on its support, the median must be located at $b_1^c = \frac{1}{2}$, and hence, $F(\frac{1}{2}) = \frac{1}{2}$, and therefore $b_{FT} = \frac{1}{2}$.

**Corollary 1.** If $F$ is symmetric in addition to strictly unimodal, $b_{FT} = \frac{1}{2}$. 

16
**Example.** Consider the incentives faced by the mayor with a 60% chance that a minimum wage increase will make the representative voter better off. Because of the positive research on the mayor’s policy idea, voters believe the mayor to be competent with a 60% probability, which corresponds to winning the election with a greater than 60% probability. Hence, actually raising the minimum wage has modest potential benefits in the event the voter likes the policy, but disastrous costs if the voter dislikes the policy. The mayor prefers to keep her current probability of re-election with certainty, rather than gamble by raising the minimum wage.

If the report by the local university had instead said that the mayor’s plan only had a 30% chance of working, then voters believe the mayor to be competent with only a 30% probability. This signal corresponds to a less than 30% chance of re-election. The mayor knows that in the likely event that the wage increase harms the voter, she will lose the election for sure. But in the unlikely event that the wage increase is beneficial, the mayor will be virtually certain to win. The mayor would prefer to gamble by raising the minimum wage.

5 **Optimal Experimentation**

If the policy choice did not have any effect on the voter’s utility in the first period, then the voter would prefer that all candidates implement their ideal policies. When the voter has only the signal $b_0$ with which to make his decision, there are incompetent incumbents who are retained when they should have been replaced by a challenger, and competent incumbents who are replaced when they should have been retained. The information from the policy experiment eliminates these mistakes, leading to higher second period welfare. If an incumbent is sufficiently reputable, then the expected policy outcome from her ideal policy in the first period is also better than the status quo, and there is no conflict between
selection and current period policy outcomes. However, this need not be the case for all incumbents. If $b_0$ is low, then the high probability of a bad policy outcome in the current period may outweigh the value of the information generated. Hence, voters face a trade-off between information gathering and current period policy outcomes. In this section, we analyze this trade-off and determine the optimal experimentation behavior.

Consider an incumbent with reputation $b_0$, and particular realization of $b^c_1 \in (0, 1)$. First, consider the case when the incumbent chooses $x_1 = s_1$. If this incumbent is retained, then the expected voter utility in the second period is $\mathbb{E}(v_2|x_1 = s, \text{retain}, b^c_1) = 2b_0 - 1$. If the incumbent is replaced by a challenger with $b^e_1$, then the expected voter utility in the second period is $\mathbb{E}(v_2|x_1 = s, \text{replace}, b^e_1) = 2b^e_1 - 1$. Now consider an incumbent with reputation $b_0$ who chooses $x_1 = \theta_1$. There is a probability $b_0$ that she is competent, and is therefore retained, and delivers a payoff of $\lambda$ in the second period. There is a probability $1 - b_0$ that she is incompetent, and are therefore replaced. The challenger delivers a payoff of $(2b^e_1 - 1)\lambda$ in the second period. Therefore, expected voter utility when $x_1 = \theta_1$ is:

$$\mathbb{E}(v_2|x_1 = \theta, b^c_1) = b_0\lambda + (1 - b_0)(2b^c_1 - 1)\lambda.$$ 

Note that in the event policy is chosen, the re-election decision never depends on the particular value of $b^c_1$, and so there is no need to condition on retention or rejection. Further, $\mathbb{E}(v_2|x_1 = \theta, b^c_1)$ is strictly greater for every realization of $b^c_1$ than when $x_1 = s$, regardless of whether incumbent is being retained or replaced. First, observe that

$$\mathbb{E}(v_2|x_1 = \theta, b^c_1) > E(v_2|x_1 = s, \text{retain}, b^c_1)$$

whenever

$$b_0\lambda + (1 - b_0)(2b^c_1 - 1)\lambda > (2b^e - 1)\lambda.$$
This is equivalent to $b_0 \lambda > b_0 (2b_1 - 1)\lambda$, which further simplifies to $1 > (2b_1 - 1)$. This inequality must hold because $b_1 > 1$.

In the case of rejection, note that

$$\mathbb{E}(v_2|x_1 = \theta, b_1^c) > E(v_2|x_1 = s, \text{ reject, } b_1^c)$$

whenever

$$b_0 \lambda + (1 - b_0)(2b_1^c - 1)\lambda > (2b_0 - 1)\lambda.$$ 

This is equivalent to $(1 - b_0)(2b_1^c - 1)\lambda > (b_0 - 1)\lambda$, which further simplifies to $(2b_1^c - 1) > -1$. This inequality must hold because $b_1^c > 0$.

We now take the expectation over all possible realizations of $b_1^c$. If $b_1^c \leq b_0$, then the incumbent will be retained if $x_1 = s$, so we integrate both sides of inequality $\mathbb{E}(v_2|x_1 = \theta, b_1^c) > E(v_2|x_1 = s, \text{ retain, } b_1^c)$. Thus,

$$\int_{0}^{b_0} \left[ b_0 \lambda + (1 - b_0)(2b_1^c - 1)\lambda \right] f(b_1^c) db_1^c > \int_{0}^{b_0} (2b_0 - 1)\lambda f(b_1^c) db_1^c.$$ 

If $b_1^c > b_0$, then the incumbent will be replaced, so we integrate both sides of inequality $\mathbb{E}(v_2|x_1 = \theta, b_1^c) > E(v_2|x_1 = s, \text{ replace, } b_1^c)$. Thus,

$$\int_{b_0}^{1} \left[ b_0 \lambda + (1 - b_0)(2b_1^c - 1)\lambda \right] f(b_1^c) db_1^c > \int_{b_0}^{1} (2b_1^c - 1)\lambda f(b_1^c) db_1^c.$$ 

Summing the two integral inequalities yields

$$\int_{0}^{1} \left[ b_0 + (1 - b_0)(2b_1^c - 1) \right] f(b_1^c) db_1^c > \int_{0}^{b_0} (2b_0 - 1)\lambda f(b_1^c) db_1^c + \int_{b_0}^{1} (2b_1^c - 1)\lambda f(b_1^c) db_1^c,$$ 

19
or equivalently,

\[ b_0 \lambda + (1 - b_0)(2E(b_1^c) - 1)\lambda > (2b_0 - 1)\lambda F(b_0) + (1 - F(b_0))\lambda[2E(b_1^c|b_1^c > b_0) - 1]. \]

The left is expected second period voter utility when the incumbent implements her ideal policy, \( E(v_2|x_1 = \theta_1) \). The right hand side is the expected second period voter utility when the incumbent chooses the status quo, \( E(v_2|x_1 = s_1) \). Hence, we have shown that:

\[ E(v_2|x_1 = \theta_1) > E(v_2|x_1 = s_1) \]

The selection effect of policy implementation is therefore the difference in expected voter utility when the incumbent implements policy, relative to when she does not, or formally,

\[ S(b_0) \equiv E(v_2|x_1 = \theta_1) - E(v_2|x_1 = s_1), \]

By the preceding arguments, we have shown that \( S(b_0) > 0 \) for every \( b_0 \in (0, 1) \).

**Lemma 2.** The selection effect, \( S(b_0) \), is positive for every \( b_0 \in (0, 1) \).

We now add the expected first period outcome, \( 2b_0 - 1 \), to arrive at the net benefit of policy implementation,

\[ NB(b_0) = S(b_0) + 2b_0 - 1. \]

It is optimal for an incumbent with reputation \( b_0 \) to implement her ideal policy if and only if \( NB(b_0) > 0 \). We now show that this will be true of any incumbent for which the expected first period policy outcome is positive.

**Lemma 3.** Assume that \( b^c, b_0 \in (0, 1) \). If \( b_0 \geq \frac{1}{2} \), then \( NB(b_0) > 0 \).
Proof. By Lemma 2, $S(b_0) > 0$. Further, $b_0 \geq \frac{1}{2}$ implies $2b_0 - 1 \geq 0$. Therefore, if $b_0 \geq \frac{1}{2}$, then $N(b_0)$ is the sum of a positive number and nonnegative number, and is therefore positive. \qed

Lemma 3 implies that the full transparency outcome is not efficient, as welfare would increase if all incumbents with $b_0 \geq \frac{1}{2}$ were to implement their ideal policies. However, Lemma 3 does not specify if it is necessarily optimal for those with $b_0 < \frac{1}{2}$ to choose to implement their ideal policies or not. It may be the case that the second period decision is more important than the first period decision, so that the costs of bad policy in the first period are less than the benefits of better selection. What we show in Theorem 2 is that it cannot ever be optimal to have only those with relatively low values of $b_0$ implementing their ideal policies.

**Theorem 2.** There exists a critical value of $b_0$, denoted $b_{opt}$, such that the net benefit of policy implementation $NB(b_0)$ is positive for $b_0 > b_{opt}$ and negative for $b_0 < b_{opt}$.

**Proof.** First, observe that $NB(0) = S(0) - 1 = -1$, because $S(0) = 0$, and

$$NB(1/2) > 0$$

by Lemma 3. Hence, because $NB(b_0)$ is continuous, there must exist at least one point, $b_{opt} \in (0, \frac{1}{2})$, such that $NB(b_{opt}) = 0$.

We now prove that there is at most one such point. Suppose there is any other point $\hat{b} \neq b_{opt}$ for which $NB(b_0) = 0$. Then by Rolles’ Theorem, there must exist at least two stationary points of $NB(b_0)$. Recall the definition of $NB(b_0)$,

$$NB(b_0) = b_0\lambda + (1-b_0)(2\mathbb{E}(b^c_1)-1)\lambda - (2b_0-1)\lambda F(b_0)-(1-F(b_0))\lambda [2\mathbb{E}(b^c_1|b^c_1 > b_0)-1]+2b_0-1,$$
which is equivalent to

\[ NB(b_0) = b_0 \lambda + (1-b_0)(2\mathbb{E}(b_1^c)-1)\lambda - (2b_0-1)\lambda F(b_0) - 2\lambda \int_{b_0}^1 b^c f(b^c)db^c + (1-F(b_0)) + 2b_0 - 1. \]

Differentiating this expression with respect to \( b_0 \) gives

\[ NB'(b_0) = \lambda - (2\mathbb{E}(b_1^c) - 1)\lambda - 2\lambda F(b_0) - 2b_0 \lambda f(b_0) + 2b_0 \lambda f(b_0) + 2 \]

Which simplifies to

\[ NB'(b_0) = \lambda - (2\mathbb{E}(b_1^c) - 1)\lambda - 2\lambda F(b_0) + 2, \]

or

\[ NB'(b_0) = 2\lambda (1 - \mathbb{E}(b_1^c) - F(b_0)) + 2. \]

Therefore, \( NB'(b_0) \) is strictly decreasing in \( b_0 \) because \( F(b_0) \) is strictly increasing in \( b_0 \). Therefore there is at most one point for which \( NB'(b_0) = 0 \), contradicting the existence of multiple roots of \( NB(b_0) \). Hence, there is at most one \( b_{opt} \). Further, because \( NB(0) < 0 \), \( NB(1/2) > 0 \), and the only root is \( b_{opt} \in (0, 1/2) \), it must be that \( NB(b_0) \leq 0 \) for all \( b_0 \leq b_{opt} \) and \( NB(b_0) > 0 \) for all \( b_0 > b_{opt} \).

Theorem 2 demonstrates the incompatibility of rational, office seeking behavior with optimal policy implementation. The incumbent seeking re-election wants to implement her ideal policy if and only if her reputation is sufficiently low. Optimal experimentation requires the incumbent to implement her ideal policy if and only if her reputation if sufficiently high.

The particular value of \( b_{opt} \) is a decreasing function of \( \lambda \). Intuitively, the more important is the second period decision relative to the first, the more valuable it is to learn the incumbent’s type, and hence, it is welfare enhancing to experiment at lower values of \( b_0 \), despite the lower
expected value of $y_1$.

**Corollary 2.** The threshold $b_{opt}$ is strictly decreasing and continuous in $\lambda$. Further, $b_{opt}(0) = \frac{1}{2}$, while $b_{opt}(\lambda) \to 0$ as $\lambda \to \infty$.

**Proof.** Differentiating $NB(b_0)$ with respect to $\lambda$ yields

$$\frac{\partial NB}{\partial \lambda} = b_0 \lambda + (1 - b_0)(2\mathbb{E}(b_1^c) - 1) - (2b_0 - 1)\lambda F(b_0) - (1 - F(b_0))\lambda[2\mathbb{E}(b_1^c|b_1^c > b_0) - 1],$$

which is always equivalent to $S(b_0)$ when $\lambda = 1$. Because $b_{opt} > 0$, we know $S(b_{opt}) > 0$, so this derivative is strictly positive. Because $NB(b_0) < 0$ for every $b_0 < b_{opt}$ and $NB(b_0) > 0$ for every $b_0 > b_{opt}$, it must be the case that

$$\frac{\partial NB}{\partial b_0} \bigg|_{b_0 = b_{opt}} > 0$$

Hence, by the Implicit Function Theorem, $b_{opt}$ is a continuous, decreasing function of $\lambda$.

Note that $NB(\frac{1}{2}, 0) = 0$, so if $\lambda = 0$, then $\bar{b} = \frac{1}{2}$. Next, note that $NB(b_{opt}) \equiv 0$ implies that $S(b_{opt}) = 1 - 2b_{opt}$, or equivalently $\lambda S(b_{opt}) \bigg|_{\lambda=1} = 2b_{opt} - 1$. Dividing on both sides gives:

$$S(b_{opt}) \bigg|_{\lambda=1} = \frac{1 - 2b_{opt}}{\lambda}$$

Taking the limit as $\lambda \to \infty$ on both sides, we find that

$$S(b_{opt}) \bigg|_{\lambda=1} = 0$$

This cannot hold for any $b_{opt} > 0$, whereas $S(0) \bigg|_{\lambda=1} = 0$. Hence, in the limit, $b_{opt} \to 0$. \(\square\)

The full transparency outcome is welfare minimizing if $b_{FT} = b_{opt}$. In that case, all reputational types with $NB(b_0) < 0$ implement their ideal policies and all reputational
types with $NB(b_0) > 0$ choose the status quo. Corollary 2 implies that if $b_{FT} < \frac{1}{2}$, then there exists $\lambda$ such that $b_{FT} = b_{opt}$. This would be the case if $F$ has a rightward skew, so that the median is less than $\frac{1}{2}$.

**Example.** Suppose that the voters of the city could force the mayor to implement the minimum wage increase, or to leave the current minimum wage as is, based on the publicly available information. In any situation where voters thought the minimum wage would work out, they would of course prefer that it be passed. They would benefit from both a better economy and they would learn the type of their mayor. For a policy that was somewhat more likely to fail than to work, they might still prefer that the mayor raise the wage so that they could learn his type. However, when it is already clear that the wage increase will fail, they would prefer that the mayor leave the wage alone, because it will likely fail and they are already quite confident that the mayor is incompetent. There is not enough value to knowing her type for sure to bear the policy costs now.

6 Private Information

In this section, we assume that the signal of the policy’s outcomes is the incumbent’s private information. When considering the issue of which policy among several to implement, private information has been shown to encourage the incumbent to act irresponsibly. The incumbent wants to implement the policy that generates favorable beliefs, and in doing so, throws away valuable private information. Full transparency would eliminate any possibility of the incumbent misrepresenting her signal to the voters through her actions. Here, we consider a choice between implementing policy and choosing the status quo, rather than multiple policies. In this environment, full transparency is detrimental because a incumbent with a good signal is able to share that signal with voters, and hence secure a high probability of re-election without doing anything. She therefore has no incentive to take a gamble by
implementing policy. If there is no way for the incumbent to credibly share her signal, then she must prove that she has a good policy idea by implementing it. Voters benefit both from more information about the competence of the incumbent, and policies that are beneficial to the voter on average.

6.1 A Baseline Private Information Setting

To formalize this argument, consider first the voter. His voting decision will be based on his expected belief about the incumbent and the challenger, due to the linearity of $\mathbb{E}(v_2|b_1)$ in $b_1$. If the incumbent implements her ideal policy, it remains fully informative, so that $b_1 = 1$ if $v_1(\theta_1) = 1$ or $b_1 = 0$ if $v_1(\theta_1) = -1$. Should the incumbent choose the status quo, the voter will form an expectation of her signal based upon the equilibrium strategy profile, equal to $\mathbb{E}(b_0|x_1 = s_1)$. The challenger will still have a publicly observable reputation of $b_1^c$, as in the full transparency case. We will also impose that $F$ is a symmetric distribution, so that $\mathbb{E}(b_1^c) = m = \frac{1}{2}$. Recall that $\mathbb{E}(b_0) = \frac{1}{2}$. An interpretation of these assumptions is that neither the incumbent nor the challenger is at any a priori advantage, and without more information, voters do not have any information if either will make things better or worse for them in the next period.

Whereas in the full transparency case, choosing the status quo implies a reputation equal to the signal of reputation, in the private information case, the status quo implies a reputation equal to the expected signal of reputation, $\mathbb{E}(b_0|x_1 = s_1)$. We refer to this as the status quo reputation for brevity. An incumbent with signal $b_0$ prefers her ideal policy when its expected probability of re-election exceeds the probability of re-election from the status quo reputation:

$$b_0 > F(\mathbb{E}(b_0|x_1 = s_1)).$$
Figure 4: The certainty equivalents for an incumbent with a low signal and an incumbent with a high signal.

Equivalently, because $F$ is strictly increasing, her ideal policy is preferred if and only if

$$F^{-1}(b_0) > \mathbb{E}(b_0 | x_1 = s_1).$$

The left hand side is the certainty equivalent of the policy implementation lottery, $CE(b_0)$. Hence, the incumbent implements her ideal policy if $CE(b_0)$ exceeds $\mathbb{E}(b| x_1 = s_1)$. She prefers the status quo if $CE(b_0) \leq \mathbb{E}(b| x_1 = s_1)$. An incumbent will only mix if $CE(b_0) = \mathbb{E}(b| x_1 = s_1)$. An example is depicted in Figure 4.

Note that $CE$ is strictly increasing in $b_0$, while $\mathbb{E}(b_0 | x_1 = s_1)$ is constant. Because the incumbent only wishes to implement her ideal policy only if $CE(b_0) \geq \mathbb{E}(b_0 | x_1 = s_1)$, if an incumbent with a reputation of $b_0$ chooses to implement her ideal policy, so will any incumbent with $b'_0 > b_0$. Second, because the incumbent chooses the status quo only if $CE(b_0) \leq \mathbb{E}(b_0 | x_1 = s_1)$, if an incumbent with a reputation of $b_0$ chooses to implement her ideal policy, so will any incumbent with $b'_0 < b_0$. Hence, the equilibrium strategies will be
Because $\mathbb{E}(b_1^e) = \frac{1}{2}$, the upper bound on $\mathbb{E}(b_0 | b_0 < b_{PI})$ is $\frac{1}{2}$. Hence, $b_{PI}$ cannot be greater than $1/2$.

characterized by a cutoff $b_{PI}$ such that any incumbent with $b_0 < b_{PI}$ chooses the status quo and any incumbent with $b_0 > b_{PI}$ chooses action.

It must be the case that $CE(b_{PI}) = \mathbb{E}(b_0 | x_1 = s_1)$. If $CE(b_{PI}) < \mathbb{E}(b_0 | x_1 = s_1)$, then there will be incumbents with $b_0 > b_{PI}$ such that $CE(b_0) < \mathbb{E}(b_0 | x_1 = s_1)$, and these types would deviate to the status quo. If $CE(b_{PI}) > \mathbb{E}(b_0 | x_1 = s_1)$, then there will be incumbents with $b_0 < b_{PI}$ such that $CE(b_0) > \mathbb{E}(b_0 | x_1 = s_1)$, and these types would deviate to action.

Due to the equilibrium strategies, it must be that $\mathbb{E}(b_0 | b_0 \leq b_{PI}) = \mathbb{E}(b_0 | b_0 \leq b_{PI})$, which is strictly increasing in $b_{PI}$. If $b_{PI} = 1$, then $\mathbb{E}(b_0 | b_0 \leq 1) = \mathbb{E}(b_0) = \frac{1}{2}$. We now show that $b_{PI} \leq \frac{1}{2}$. Suppose, contrary to our claim, that $b_{PI} > \frac{1}{2}$. For every $b_0 \in (\frac{1}{2}, b_{FT})$, it must be the case that $CE(b_0) > \frac{1}{2} > \mathbb{E}(b_0 | b_0 \leq b_{PI})$, so these incumbents would prefer to deviate to action. This is illustrated in Figure 5.

Further, it cannot be the case that $b_{PI} \in (0, \frac{1}{2}]$. Suppose it were. Note that $\mathbb{E}(b_0 | b_0 \leq b_{PI}) < b_{PI}$, and that for every $b_0 \in (0, \frac{1}{2})$, the certainty equivalent $CE(b_0)$ is strictly greater than $b_0$, as shown in Figure 6.

Hence, $\mathbb{E}(b_0 | b_0 \leq b_{PI}) < CE(b_{PI})$, contrary to the requirement that $CE(b_{PI}) = \mathbb{E}(b_0 | b_0 \leq b_{PI})$. The only possible value for $b_{PI}$ is therefore zero.
Figure 6: For $b_{PI} < \frac{1}{2}$, $CE(b_{PI}) > b_{PI}$, whereas $E(b_0|b_0 < b_{PI}) < b_{PI}$, for any $b_{PI} > 0$. The blue line is $y = b_{PI}$.

If $b_{PI} = 0$, then the only type who ever chooses the status quo with positive probability is $b_0 = 0$. Hence, $E(b_0|b_0 \leq b_{PI}) = 0$. Further, $CE(0) = 0$, and so $E(b_0|x_1 = s_1) = CE(b_{PI})$ at $b_{PI} = 0$. Therefore, $b_{PI} = 0$ is the unique equilibrium cutoff.

The equilibrium strategy under this private information model is given formally in Theorem 3.

**Theorem 3.** When the incumbent has a private signal, $E(b_0) = E(b_1) = \frac{1}{2}$, and $F$ is unimodal, all incumbents implement their ideal policies.

Under the assumptions that $F$ is unimodal and symmetric, and that $E(b_1) = \frac{1}{2}$, the private information equilibrium generates strictly higher expected welfare for the voter. To see this, note that for any $b_0 < \frac{1}{2}$, there is no change in equilibrium behavior, and hence, these reputational types do not contribute to any difference in welfare. For $b_0 > \frac{1}{2}$, these reputational types are implementing their ideal policies in the private information case, whereas they were choosing the status quo in the full information case. By Lemma 3, it is welfare enhancing to have types with $b_0 > \frac{1}{2}$ implement her ideal policy, and so the private information equilibrium is better for voters than full transparency.
The intuition is as follows. When the signal is public, incumbents with a high signal are able to fully capture the electoral benefits of having that signal without incurring the risk of actually implementing their ideal policies. When the signal is private, the only way for the incumbent to benefit from having a high signal is to actually implement her ideal policy. If she chooses the status quo, then the content of her signal is irrelevant to her re-election prospects. Hence, high signal incumbents are now encouraged to actually implement their ideal policies, and this raises welfare for voters.

In the appendix, we show that the equilibrium in Theorem 3 also exists when $F$ is not necessarily symmetric, the expected second period outcome given reputation, $E(v_2|b_0)$, is any strictly increasing function, and the experiment may not be fully informative.

### 6.2 An Alternative Private Information Model

In the first private information scenario, all types implemented policy, because the status quo reputation is always sufficiently low to make any incumbent with $b_0 > 0$ prefer the gamble of implementation. We now consider the possibility of an equilibrium in which low reputational types are discouraged from implementing policy as well.

Suppose that with probability $\gamma$, when an incumbent chooses to implement her ideal policy, it is not be implemented before the next election, and that voters do not observe the action. Hence, with probability $\gamma$, when $x_1 = \theta_1$, it instead appears that $x_1 = s_1$. An interpretation is that sometimes, a policy won’t be implemented in time for the next election, and voters do not keep abreast of what the incumbent is trying to do; they only observe the policies that are actually carried out.

In this informational environment, when the voter observes the status quo, he does not know if this is because the incumbent tried to implement her policy and it failed to happen, or if the incumbent deliberately chose to keep the status quo. This informational change
increases the status quo reputation, because for any given cutoff $b_{\text{cut}}$, voters know that with probability $\gamma$, the incumbent actually had a signal greater than $b_{\text{cut}}$. The higher status quo reputation induces some low reputational types to instead choose the status quo rather than their ideal policy.

As $\gamma$ increases, voters are spared from some bad policy outcomes in the first period, which is welfare improving. Unfortunately, voters also lose some information about those types, and this may worsen selection. We derive conditions under which this alternative private information setting has higher welfare than the fully transparent case.

First, we show the behavior of incumbents in this informational environment. An incumbent wants to implement her ideal policy if and only if:

$$(1 - \gamma)b_0 + \gamma F(\mathbb{E}(b_0|x_1 = s_1)) \geq F(\mathbb{E}(b_0|x_1 = s_1))$$

Which simplifies algebraically to

$$b_0 \geq F(\mathbb{E}(b_0|x_1 = s_1)).$$

Applying $F^{-1}$ to both sides, we obtain

$$CE(b_0) \geq \mathbb{E}(b_0|x_1 = s_1)$$

Hence, as in the baseline case, incumbents want to implement her ideal policy if and only if $b_0$ is sufficiently high, implying a cutoff $b_{\text{alt}}$ characterizes the equilibrium. For an arbitrary cutoff $b_{\text{cut}}$, the status quo is chosen by any incumbent with $b_0 \leq b_{\text{cut}}$ and occurs involuntarily with probability $\gamma$ for any incumbent with $b_0 > b_{\text{cut}}$. This implies

$$\mathbb{E}(b_0|x_1 = s_1) = \frac{\int_0^{b_{\text{cut}}} g(b_0)b_0 db_0 + \gamma \int_{b_{\text{cut}}}^1 g(b_0)b_0 db_0}{G(b_{\text{cut}}) + \gamma(1 - G(b_{\text{cut}}))}.$$
Note that at $b_{\text{cut}} = 0$, this is:

$$\frac{\gamma \int_0^1 g(b_0)b_0 \, db_0}{G(0) + \gamma (1 - G(0))} = \frac{\int_0^1 g(b_0)b_0 \, db_0}{1} = \mathbb{E}(b_0) = \frac{1}{2}.$$ 

Unlike the first private information model, we cannot have $b_{\text{alt}} = 0$, because $CE(0) = 0 \neq \frac{1}{2}$.

To determine where a solution may be located, we first consider the derivative of $\mathbb{E}(b_0|x_1 = s_1)$ with respect to $b_{\text{cut}}$. Differentiation implies

$$\frac{d\mathbb{E}(b_0|x_1 = s_1)}{db_{\text{cut}}} = \frac{[g(b_{\text{cut}})b_{\text{cut}} - \gamma g(b_{\text{cut}})b_{\text{cut}}](F(b_{\text{cut}}) + \gamma (1 - F(b_{\text{cut}}))) - \int_0^{b_{\text{cut}}} g(b_0)b_0 \, db_0 + \gamma \int_{b_{\text{cut}}}^1 g(b_0)b_0 \, db_0}{(G(b_{\text{cut}}) + \gamma (1 - G(b_{\text{cut}))))^2},$$

Which simplifies to

$$\frac{d\mathbb{E}(b_0|x_1 = s_1)}{db_{\text{cut}}} = \frac{[g(b_{\text{cut}})b_{\text{cut}} - \gamma g(b_{\text{cut}})b_{\text{cut}}] - \mathbb{E}(b_0|x_1 = s_1)(g(b_{\text{cut}}) - \gamma g(b_{\text{cut}}))}{(G(b_{\text{cut}}) + \gamma (1 - G(b_{\text{cut}))))},$$

and further to

$$\frac{d\mathbb{E}(b_0|x_1 = s_1)}{db_{\text{cut}}} = \frac{(1 - \gamma)g(b_{\text{cut}})(b_{\text{cut}} - \mathbb{E}(b_0|x_1 = s_1))}{(G(b_{\text{cut}}) + \gamma (1 - G(b_{\text{cut}))))}.$$ 

Hence, this function is decreasing, stationary, or increasing as $(b_{\text{cut}} - \mathbb{E}(b_0|x_1 = s_1))$ is less than, equal to, or greater than 0.

To further characterize $\mathbb{E}(b_0|x_1 = s_1)$, let $\Delta(b_{\text{cut}}) \equiv b_{\text{cut}} - \mathbb{E}(b|x_1 = s_1)$.

**Lemma 4.** The function $\Delta(b_{\text{cut}})$ has a unique point $\tilde{b}$ for which $\Delta(\tilde{b}) = 0$. The conditional expectation $\mathbb{E}(b_0|x_1 = s_1)$ is strictly decreasing on $(0, \tilde{b})$ and strictly increasing on $(\tilde{b}, 1)$.

**Proof.** Observe that $\Delta(0) < 0$ and $\Delta(1) > 0$. Hence, there must exist at least one point $b_{PI1}$ such that $\Delta(\tilde{b}) = 0$ by the Intermediate Value Theorem.

Suppose there existed multiple solutions to $\Delta(\tilde{b}) = 0$. Order these solutions from least to greatest as $\tilde{b}_1, \tilde{b}_2, ..., \tilde{b}_n$. Differentiating reveals that $\Delta'(b_{\text{cut}}) = 1 - \frac{d\mathbb{E}(b_0|x_1 = s_1)}{db_{\text{cut}}}$. Thus, at any
such that $\Delta(\tilde{b}) = 0$, we know that $\frac{dE(b|x_1=s_1)}{db_{cut}} = 0$, so $\Delta'(\tilde{b}_1) = 1$. Therefore by Lemma 9, $\Delta(b_{cut}) > 0$ for any $b_{cut} \in (\tilde{b}_1, \tilde{b}_2)$. However, $\Delta'(\tilde{b}_2) = 1$, so for $b_{cut} < \tilde{b}_2$ but sufficiently close, $\Delta(b_2) < 0$, contradicting the existence of $\tilde{b}_2$. Hence, $\tilde{b}$ is unique.

Therefore, $\Delta(b_{cut}) < 0$ for all $b_{cut} < \tilde{b}$, and $\Delta(b_{cut}) > 0$ for all $b_{cut} > \tilde{b}$. Consequently, $\frac{dE(b|x_1=s_1)}{db_{cut}} < 0$ for all $b_{cut} < b_{cut1}$, and $\frac{dE(b|x_1=s_1)}{db_{cut}} > 0$ for all $b_{cut} > \tilde{b}$. 

\begin{lemma}
There is a unique $b_{alt}$, located in $(0, \frac{1}{2})$, such that $E(b|x_1 = s_1) = CE(b_{alt})$.
\end{lemma}

\begin{proof}
At $b_{cut} = 0$, we know that $CE(b_{cut}) - E(b_0|x_1 = s_1) < 0$. Further, because $CE(b_{cut}) > b_{cut}$ on $(0, \frac{1}{2})$, we know that at $\tilde{b}$, $CE(b_{cut}) - E(b_0|x_1 = s_1) = CE(b_{cut}) - b_{cut} > 0$. Hence, by the Intermediate Value Theorem, there exists a $b_{alt} \in (0, \tilde{b})$ for which $CE(b_{alt}) - E(b_0|x_1 = s_1) = 0$.

Because $E(b_0|x_1 = s_1) = \frac{1}{2}$ at $b_{cut} = 0$ and $\frac{dE(b_0|x_1=s_1)}{db_{cut}} < 0$ whenever $b_{cut} < \tilde{b}$, it it must be the case that at $b_{alt}$, $E(b_0|x_1 = s_1) < \frac{1}{2}$. This implies that $b_{alt} < \frac{1}{2}$.

Further, because $\tilde{b}$ is a unique root of $\Delta(b_{cut})$, and $\Delta(b_{cut}) > 0$ for every $b_{cut}$ greater than $\tilde{b}$ by Lemma 4, it must be the case that $b_{cut} > E(b_0|x_1 = s_1)$. Because $CE(b_{cut}) > b_{cut}$ for $b_{cut} \in (0, \frac{1}{2})$, there is no possibility of a second point for which $E(b_0|x_1 = s_1) = CE(b_{cut})$ in $(b_{cut}, \frac{1}{2})$. Hence, there is a unique equilibrium value of $b_{alt}$, and it must lie in $(0, \frac{1}{2})$.

Because $CE(b_0) < E(b_0|x_1 = s_1)$ for every $b_0 < b_{alt}$, we know all such incumbents prefer the status quo. Because $CE(b_0) > E(b_0|x_1 = s_1)$ for every $b_0 > b_{alt}$, we know all such incumbents prefer to implement their ideal policies. Hence, the following summarizes the equilibrium of this private information model.

\begin{theorem}
If voters mistakenly observe $x_1 = s_1$ with probability $\gamma$ when $x_1 = \theta_1$, then there is a unique value $b_{alt} \in (0, \frac{1}{2})$ such that incumbents choose $x_1 = \theta_1$ if $b_0 \geq b_{alt}$, and such that incumbents choose $x_1 = s_1$ if $b_0 < b_{alt}$.
\end{theorem}

The welfare comparison of this to full information is a work in progress. It is certainly an improvement in first period welfare. However, the loss of information about the incumbents
implies a cost in terms of worse second term policy. Voters are forced to use \( \mathbb{E}(b_0|x_1 = s_1) \)
rather than the actual value of \( b_0 \) in making decisions about any incumbent with \( b_0 < b_{\text{alt}} \)
and for \( \gamma \) percent of those with higher \( b_0 \). Provided \( \gamma \) is not too large, then not too much
information is lost, and it should be the case that the second period costs are smaller than
the first period benefits.

**Example.** Suppose that the city government’s impact study on the effects of raising the
minimum wage from 9 to 15 is sent only to the mayor. Voters know that the mayor would
want to implement the policy only if it is sufficiently likely to work. Hence, when voters
observe a mayor leaving the minimum wage at 9 dollars per hour, they believe that the
report must have indicated something bad- suppose they know \( b_{\text{alt}} = 15\% \). They don’t know
for sure what the report said, but they can infer an average, and use it to make their voting
decision. The mayor with a 60\% chance of improving the local economy would prefer to
raise the minimum wage, because a 60\% chance of re-election is better than the relatively
low chance she has by leaving the wage alone. A mayor with a 10 \% chance of improving
the economy decides that it would be better to accept the modest probability of re-election
offered by keeping the status quo than to gamble. Voters now get policies enacted when they
are likely to work, and the policies least likely to work are no longer passed. In addition,
voters are learning more about their mayor, meaning they can make a better electoral choice.

7 Conclusion

We have shown that under full transparency, the incumbent’s risk attitudes at different levels
of reputation cause those who should implement their ideal policy to choose the status quo,
and those who should choose the status quo to instead implement their ideal policies. Despite
having no private information, incumbents use their ability to control the flow of information
to improve their expected probability of re-election. When information is private, because
there is a common status quo reputation, the incumbent’s payoff from the status quo is no longer tied to her signal. For those with high signals, the status quo reputation is low enough to encourage them to gamble on implementing their ideal policies, while those with low beliefs prefer to choose the status quo and get a relatively high status quo reputation. As we show in Section 8, these results do not depend upon a fully informative experiment or perfect Bayesian updating by voters, the uncertainty about the challenger, or the linearity of future policy outcomes in $b_0$. Instead, the fundamental cause of the incumbent’s behavior is the endogenous risk preference, induced by the unimodal CDF of challengers.

One could also consider the case when incumbents have two dimensions on which to signal- competence and motivation. For example, if it is understood by voters that incumbents would prefer not to implement policy when they have a high reputation, then the decision to implement policy anyway could potentially serve as a signal of policy motivation. If motivations were publicly known as well, this incentive would not exist. Similarly to our model, to the extent that “bad” types are able to signal their way to re-election, this implies worse second period outcomes, and a similar tradeoff of current against future policy utility.

Our findings help to explain why institutions may be set up to have private information in the first place, given the incentives for pandering and posturing private information may induce. Incentive correcting mechanisms such as options, which are available in the private sector, are not available in political settings. Therefore, allowing incumbents to have private information may be one of the only tools available to encourage more responsible risk-taking.
8 Appendix

8.1 Generalization of Full Transparency Equilibrium

In this subsection, we consider the case when the lottery the incumbent faces is not fully informative of her type, and may not even be the result of Bayesian updating. Instead, we will allow for a generic good and bad outcome, so long as they meet certain properties also satisfied by Bayesian updating.

Formally, let $b_1 = b_+(b_0)$ be the the belief of the voter that the incumbent is competent, given $v_1 = 1$, and $b_1 = b_-(b_0)$ be the belief of the voter that the incumbent is incompetent, conditional on $v_1 = -1$. Finally, let $b_1 = b_0$ if $v_1 = 0$. We require that $b_+(b_0)$ and $b_-(b_0)$ are both strictly increasing and continuous in $b_0$, and $0 \leq b_-(b_0) < b_0 < b_+(b_0) \leq 1$, $\forall b_0 \in (0, 1)$.

This is satisfied by Bayesian updating when there is a positive probability that incompetent incumbents have success at times and that competent incumbents sometimes fail. However, it could also be satisfied by any other process voters might use to update beliefs, as long as more positive priors map into more positive beliefs after seeing the result of policy, and once the voter is certain about the type of the incumbent, no more learning can happen.

We maintain the assumptions that $F$ is unimodal and that $f(0) = 0$ and $f(1) = 0$. The latter restriction eliminates the possibility of kinks in $D$, enabling a much easier characterization of its properties.

Define $D(b_0; b_-, b_+) \equiv \frac{b_0 - b_-}{b_+ - b_-}F(b_+) + (1 - \frac{b_0 - b_-}{b_+ - b_-})F(b_-) - F(b_0)$. This is the utility differential from implementing her ideal policy relative to the status quo, given an exogenous $b_-$ and $b_+$. Because we will only differentiate with respect to and evaluate at varying values of $b_0$, we will write $D(b_0) \equiv D(b_0; b_-, b_+)$ as long as $b_-$ and $b_+$ remain fixed. As in the main text, the incumbent only prefers to implement her ideal policy if $D(b_0) \geq 0$, and prefers the status quo if $D(b_0) \leq 0$. We begin our analysis by deriving properties of $D$ given fixed values for $b_+$ and $b_-$. 
Lemma 6. There exist values $b'$ and $b''$, with $b' < b''$ such that

(i) $D'(b_0) < 0$ whenever $b_0 \in (b', b'')$

(ii) $D'(b_0) > 0$ whenever $b_0 \notin (b', b'')$

(iii) $D'(b_0) = 0$, whenever $b_0 \in \{b', b''\}$

Proof. Differentiation of $D$ implies

$$D'(b_0) = \frac{F(b_+ - F(b_-)}{b_+ - b_-} - f(b_0)$$

Hence, $D'(b_0)$ is positive if and only if $\frac{F(b_+ - F(b_-)}{b_+ - b_-} > f(b_0)$, negative if and only if $\frac{F(b_+ - F(b_-)}{b_+ - b_-} < f(b_0)$, and zero if and only if $\frac{F(b_+ - F(b_-)}{b_+ - b_-} = f(b_0)$. Note that $\frac{F(b_+ - F(b_-)}{b_+ - b_-}$ is a constant with respect to $b_0$, so this is equivalent to identifying the lower contour, upper contour, and level sets of $f$. By the strict quasiconcavity of $f$, the upper contour set must be convex, or an interval.

We first prove that the upper set is always nonempty. In order for it to be empty, it would have to be true that $\frac{F(b_+ - F(b_-)}{b_+ - b_-} \geq \max_{b_0 \in [0, 1]} f(b_0)$. Because $f$ is unimodal, the maximum is attained at $\mu$. So, supposing that $\frac{F(b_+ - F(b_-)}{b_+ - b_-} > f(\mu)$, note that this implies

$$F(b_+) - F(b_-) > f(\mu)(b_+ - b_-),$$

or equivalently

$$\int_{b_-}^{b_+} f(b_0) db_0 > \int_{b_-}^{b_+} f(\mu) db_0.$$

Since $f(\mu) > f(b_0)$ for every $b_0 \neq \mu$, this is impossible. Hence $\frac{F(b_+ - F(b_-)}{b_+ - b_-} < f(\mu)$, and therefore must have a nonempty upper set. Further, because $f$ is strictly increasing on $(0, \mu)$ and then decreasing on $(\mu, 1)$, it must be the case that $b' < \mu$ and $b'' > \mu$. Hence, for any $b_0 \in (b', b'')$, $D'(b_0) > 0$. For any $b_0 \notin [b_-, b_+]$, strict quasiconcavity implies $\frac{F(b_+ - F(b_-)}{b_+ - b_-} >
\( f(b_0) \) and so \( D'(b_0) < 0 \). Because \( f(b_0) \) is continuous, \( D'(b_0) \) is continuous, and therefore \( D(b') = D(b'') = 0 \). 

Let \( b_{FT} \) be a value of \( b_0 \) such that (i) \( D(b_{FT}) = 0 \) and (ii) either \( b_{FT} \notin \{b_-, b_+\} \) or \( D'(b_{FT}) = 0 \). The point \( b_{FT} \) is unique, and characterizes the preferences of incumbents. All incumbents with \( b_0 \leq b_{FT} \) prefer implementation and all incumbents with \( b_0 \geq b_{FT} \) prefer the status quo. The case in which there is exactly one stationary point of \( D \) is considered first.

**Lemma 7.** Given \( b_- \) and \( b_+ \), there exists \( b_{FT} \) such that all incumbents with \( b_0 \leq b_{FT} \) prefer implementation and all incumbents with \( b_0 \geq b_{FT} \) prefer the status quo.

**Proof.** Note that \( D(b_-) = D(b_+) = 0 \). By Rolle’s theorem, there must exist a stationary point in \((b_-, b_+)\). Hence, there are three cases to consider:

1. \( b' \in (b_-, b_+) \) and \( b'' \notin (b_-, b_+) \)
2. \( b' \notin (b_-, b_+) \) and \( b'' \in (b_-, b_+) \)
3. \( b' \in (b_-, b_+) \) and \( b'' \in (b_-, b_+) \)

In the first case, it must be the case that \( b'' \geq b_+ \). Therefore, it cannot be the case that \( b_{FT} < b_+ \), as there would not be a second stationary point and \( D'(b_-) \neq 0 \), because \( b_- < b' \). If \( b'' = b_+ \), then \( b_{FT} = b_+ \). It cannot also be the case that there exists \( b_{FT} > b_+ \), because there would be no stationary point in \((b_+, b_{FT})\). Hence, it must be unique. If \( b'' > b_+ \), then \( D'(b_+) < 0 \), and therefore for a point \( b > b_+ \) but arbitrarily close, \( D(b) < 0 \). Further, as \( b_0 \to \infty \), \( D(b_0) \to \infty \). Hence, by the Intermediate Value Theorem, there must exist \( b_{FT} \in (b_+, \infty) \) such that \( D(b_{FT}) = 0 \). Since there are only two stationary points, there can only be one such \( b_{FT} \), and neither \( D'(b_+) \) nor \( D'(b_-) \) is 0. Hence, \( b_{FT} \) must be unique. Note that in this case, \( D(b_0) > 0 \) for all \( b_0 \in (b_-, b_+) \), because \( D'(b_+) < 0 \) implies that for \( b < b_+ \),
but arbitrarily close, \( D(b) > 0 \). Because \( D(b_0) \neq 0 \) for any \( b_0 \in (b_-, b_+) \), it must be that \( D(b) > 0 \) for every \( b_0 \in (b_-, b_+). \) Further, for any \( b_0 \in (b_-, b_+, b_0 < b_{FT} \)

In the second case, it must be the case that \( b' \leq b_- \). Therefore, it cannot be the case that \( b_{FT} > b_- \), because there would not be a second stationary point and \( D'(b_+) \neq 0 \) because \( b_+ > b' \). If \( b' = b_- \), then \( b_{FT} = b_- \). It cannot also be the case that there exists \( b_{FT} < b_- \), because there would be no stationary point in \((b_{FT}, b_-)\). Hence, it must be unique. If \( b' < b_- \) then \( D'(b_-) < 0 \), and therefore for a point \( b < b_- \) but arbitrarily close, \( D(b) > 0 \). Further, as \( b_0 \to -\infty \), \( D(b_0) \to -\infty \). Hence, by the Intermediate Value Theorem, there must exist \( b_{FT} \in (-\infty, b_-) \) such that \( D(b_{FT}) = 0 \). Since there are only two stationary points, there can only be one such \( b_{FT} \), and neither \( D'(b_+) \) nor \( D'(b_-) \) is 0. Hence, \( b_{FT} \) must be unique. Note that in this case, \( D(b_0) < 0 \) for all \( b_0 \in (b_-, b_+) \), because \( D'(b_-) < 0 \) implies that for \( b > b_- \) but arbitrarily close, \( D(b) < 0 \). Because \( D(b_0) \neq 0 \) for any \( b_0 \in (b_-, b_+) \), it must be that \( D(b) < 0 \) for every \( b_0 \in (b_-, b_+) \). Further, for any \( b_0 \in (b_-, b_+, b_0 > b_{FT} \)

In the third case, it must be that \( b_{FT} \in (b_-, b_+) \) by Rolles’ Theorem. Note that \( D'(b_-) > 0 \) because \( b_- < b' \), which implies for \( b > b_- \) but arbitrarily close, \( D(b) > 0 \). Similarly, \( D'(b_+) > 0 \), which implies that for \( b < b_+ \) but arbitrarily close, \( D(b) < 0 \). Hence, by the Intermediate Value Theorem, there must exist \( b_{FT} \in (b_-, b_+) \) such that \( D(b_{FT}) = 0 \). Because neither \( D'(b_+) \) nor \( D'(b_-) \) equal 0, and there are only two stationary points, there cannot be another \( b_{FT} \). Hence, it is unique.

Note that in this case, \( D(b_0) > 0 \) for all \( b_0 \in (b_-, b_{FT}) \), because \( D'(b_-) > 0 \) implies that for \( b > b_- \) but arbitrarily close, \( D(b) > 0 \). Because \( D(b_0) \neq 0 \) for any \( b_0 \in (b_-, b_{FT}) \), it must be that \( D(b) < 0 \) for every \( b_0 \in (b_-, b_{FT}) \). A symmetric argument holds for \( b_0 \in (b_{FT}, b_+) \).

Since the three cases presented are exhaustive, we have shown that \( b_{FT} \) is always defined, unique, and that for any \( b_0 \in (b_-, b_+) \), \( b_0 \leq b_{FT} \) implies a preference for implementation, and \( b_0 \geq b_{FT} \) implies a preference for the status quo. \( \square \)
What we have shown so far for an arbitrary choice of $b_-$ and $b_+$, we can define a value $b_{FT}$ that characterizes candidate preference. We now shift our attention to the specific $b_-$ and $b_+$ arising from $b_0$, and show that as $b_0$ increases, $b_{FT}$ decreases, so that there is a value $b^{**}$ such that $b_0 < b^{**}$ implies a preference for action and $b_0 > b^{**}$ implies a preference for the status quo, for the actual lottery faced by those incumbents.

Formally, we now consider

$$D(b_0, b_+(b_0), b_-(b_0)) = \frac{b_0 - b_-(b_0)}{b_+(b_0) - b_-(b_0)} F(b_+(b_0)) + \left(1 - \frac{b_0 - b_-(b_0)}{b_+(b_0) - b_-(b_0)}\right) F(b_-(b_0)) - F(b_0),$$

So that $D$ gives the utility difference between ideal policy implementation and the status quo for the lottery faced by an incumbent with reputation $b_0$. We know from Lemma (something) that there exists a unique $b_{FT}$ for this lottery such that the incumbent prefers $x_1 = \theta_1$ if and only if $b_0 \leq b_{FT}$ and $x_1 = s_1$ if and only if $b_0 \geq b_{FT}$. We show in the following lemma that $b_{FT}$ is a decreasing function of $b_0$, so that if a lottery is accepted at reputation $b_0$, it will also be accepted for any lower $b_0$, and similarly, if the status quo is preferred at any $b_0$, then it is also preferred for any greater $b_0$. Hence, there exists a $b^{**}$ that divides those types.

**Lemma 8.** There exists $b^{**}$ such that $b_0 \leq b^{**}$ implies policy implementation is preferred and $b_0 \geq b^{**}$ implies the status quo is preferred.

**Proof.** Because $b_{FT}$ for any particular $b_0$ must always satisfy $D(b_{FT}) = 0$ we will use the Implicit Function Theorem to prove that $b_{FT}$ must actually be a decreasing function of $b_0$.

Differentiation shows that

$$\frac{\partial D}{\partial b_{FT}} = \frac{F(b_+)}{b_+ - b_-} - f(b^{**}).$$

This is positive if and only if $\frac{F(b_+)}{b_+ - b_-} > f(b_{FT})$, or equivalently, if $\frac{dD}{db_0}\big|_{b_0=b_{FT}} < 0$. 

39
For $b_+$, differentiation implies

$$\frac{\partial D}{\partial b_+} = -\frac{1}{b_+ - b_-} F(b_+) + \frac{b_T - b_-}{b_+ - b_-} f(b_+) + \frac{1}{b_+ - b_-} b_T - b_- F(b_-),$$

Which is positive if and only if

$$-\frac{b_T - b_-}{b_+ - b_-} F(b_+) + (b_T - b_-) f(b_+) + \frac{b_T - b_-}{b_+ - b_-} F(b_-) > 0$$

which is further reduced to

$$(b_T - b_-) [f(b_+) - \frac{F(b_+) - F(b_-)}{b_+ - b_-}].$$

Note that this is also equal to

$$(b_T - b_-) [-\frac{\partial D}{db_0} |_{b_0 = b_+}].$$

This is positive if $b_T > b_- \text{ and } \frac{\partial D}{db_0} |_{b_0 = b_+} < 0$, or if $b_T < b_- \text{ and } \frac{\partial D}{db_0} |_{b_0 = b_+} > 0$.

For $b_-$, differentiation implies

$$\frac{\partial D}{\partial b_-} = \frac{b_T - b_+}{(b_+ - b_-)^2} F(b_+) + \frac{b_- - b_T}{(b_+ - b_-)^2} F(b_-) - \frac{b_- - b_T}{b_+ - b_-} F(b_-),$$

this is positive if and only if

$$\frac{b_T - b_+}{(b_+ - b_-)} F(b_+) + \frac{b_- - b_T}{(b_+ - b_-)} F(b_-) + (b_+ - b_T) f(b_-) > 0$$

and

$$-\frac{b_T - b_-}{(b_+ - b_-)} F(b_+) + \frac{b_- - b_T}{(b_+ - b_-)} F(b_-) + (b_+ - b_T) f(b_-) > 0$$

and

$$\frac{b_T - b_+}{(b_+ - b_-)} F(b_+) - \frac{b_- - b_T}{(b_+ - b_-)} F(b_-) - (b_+ - b_T) f(b_-) < 0$$
\[ (b_+ - b_{FT}) \left[ \frac{F(b_+) - F(b_-)}{b_+ - b_-} - f(b_-) \right] < 0 \]

This is equivalent to:

\[ (b_+ - b_{FT}) \left[ \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_-} \right] < 0 \]

This is satisfied if \( b_{FT} > b_+ \) and \( \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_-} > 0 \) or if \( b_{FT} < b^+ \) and \( \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_-} < 0 \).

In any situation in which \( \frac{\partial D}{\partial b_+} > 0, \frac{\partial D}{\partial b_-} > 0, \) and \( \frac{\partial D}{\partial b_{FT}} > 0, \) an increase in \( b_0 \) which thereby increases \( b_+ \) and \( b_- \) will necessarily decrease \( b_{FT} \). We proceed by considering each possible case from Lemma (something).

In case (i), \( b_{FT} \geq b_+ > b_- \), \( \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_+} < 0 \), and \( \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_-} > 0 \). Further, because \( b_{FT} < b' \), \( \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_{FT}} > 0 \). Hence, by the Implicit Function Theorem, \( \frac{db_{FT}}{db_0} \leq 0 \), and it varies continuously with \( b_0 \).

In case (ii), \( b_{FT} \leq b_- < b_+ \), \( \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_+} > 0 \), and \( \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_-} < 0 \). Further, because \( b_{FT} > b'' \), \( \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_{FT}} > 0 \) and so by the Implicit Function Theorem, \( \frac{db_{FT}}{db_0} \leq 0 \), and it varies continuously with \( b_0 \).

In case (iii), we instead have all partial derivatives of \( D \) negative, because \( b_- < b_{FT} < b_+ \), \( \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_-} > 0 \), and \( \frac{\partial D}{\partial b_0} \bigg|_{b_0 = b_+} < 0 \). Because \( b_{FT} \in (b_-, b_+) \), so \( \frac{db_{FT}}{db_0} \bigg|_{b_0 = b_{FT}} < 0 \).

Hence, \( b_{FT}(b_0) \) is a non-increasing, continuous function of \( b_0 \). There are therefore three possibilities for incumbent behavior:

1. \( b_{FT}(0) \leq 0 \). In this case, all incumbents prefer the status quo, and \( b^{**} = 0 \).

2. \( b_{FT}(1) \geq 1 \). In this case, all incumbents prefer action, and \( b^{**} = 1 \).

3. \( b_{FT}(0) > 0 \) and \( b_{FT}(1) < 1 \), so there is a unique point \( b^{**} \in (0,1) \) such that \( b_{FT}(b^{**}) = b^{**} \), and for all \( b_0 < b^{**} \), it holds that \( b_0 < b_{FT}(b_0) \) and incumbents prefer to implement their ideal policies, while for all \( b_0 > b^{**} \), it holds that \( b_0 > b_{FT}(b_0) \), so those incumbents prefer the status quo.
8.1.1 Known Challenger

The case in which a challenger is known is represented by a CDF that is zero until the challenger quality, $p^c$, and then increases to one at the challenger quality, or formally,

$$F(b^c_i) = \begin{cases} 
0 & \text{if } b^c_i < p^c \\
1 & \text{if } b^c_i \geq p^c
\end{cases}$$

Given $b_0$, and a fully informative policy experiment, the incumbent only prefers implementation when $b_0 < p^c$. In that case, implementing policy implies a positive probability of winning, while she will lose for sure with the current reputation. If $b_0 > p^c$ then the incumbent will win for sure if she keeps her current reputation. If she decides to try policy, there is a risk of losing. Hence, the status quo is strictly preferred. At $b_0 = p^c$, the incumbent is indifferent. Therefore, the incumbent prefers to implement her ideal policy if and only if $b_0 \leq p^c$.

8.1.2 Allowing for Valence

Suppose that the challenger, in addition to her reputation, has a valence advantage over the incumbent $v^c \in \mathbb{R}$. Let the challenger’s quality be $q^c = b^c_i + v^c$, and let it be distributed according to strictly unimodal CDF $J$. This would be the case, for example, if both $b^c_i$ and $v^c$ were distributed normally, with the former being truncated.

Let the expected voter utility of electing a challenger with $b^c_i$ and $v^c$ be

$$\mathbb{E}(v_2|b^c_i, v^c) = \mathbb{E}(v_2|b^c_i) + 2v^c,$$

Where $\mathbb{E}(v_2|b^c_i) = 2b^c_i - 1$ as in the main text. The decision to give valence a weight of two
rather than one is purely for convenience, and without loss of generality. The incumbent is preferred by the voter if and only if:

\[ 2b_1 - 1 \geq 2b^c_1 - 1 + 2v^c \]

\[ b_1 > b^c_1 + v^c \]

\[ b_1 > q^c \]

And therefore the probability of re-election given \( b_1 \) is \( J(b_1) \). The expected probability of re-election given \( x_1 = \theta_1 \) is \( b_0J(1) + (1-b_0)J(0) \). Unlike before, \( J(1) < 1 \) and \( J(0) > 0 \), due to the effect of valence- an incumbent certain to be competent may still lose to a challenger with sufficient valence, and an incumbent certain to be incompetent may still win over a challenger with low enough valence. In the event the status quo is chosen, then the re-election probability is \( J(b_0) \). Hence, the incumbent prefers to implement her ideal policy if and only if \( b_0J(1) + (1-b_0)J(0) > J(b_0) \). Defining \( D(b_0) = b_0J(1) + (1-b_0)J(0) - J(b_0) \), the incumbent prefers to implement her ideal policy only if \( D(b_0) \geq 0 \). In this case, \( D(0) = J(0) - J(0) = 0 \) and \( D(1) = J(1) - J(1) = 0 \). Differentiating with respect to \( b_0 \), we find that

\[ D'(b_0) = J(1) - J(0) - j(b_0) \]

Hence, it must be the case that \( j(b_0) = J(1) - J(0) \) at any stationary point. As in the proof in the main text, strict unimodality of \( j \) implies at most two points where \( D'(b_0) = 0 \). Replacing the restriction that \( 1 > f(1) \) and \( 1 > f(0) \) by \( J(1) - J(0) > j(0) \) and \( J(1) - J(0) > j(1) \), the proof of Proposition 1 applies exactly.
8.2 Generalization of First Private Info Equilibrium

In this subsection, we show that the equilibrium in Theorem 3 also exists when $F$ is not necessarily symmetric, the expected second period outcome given reputation, $y_2(b_0)$, is any strictly increasing function, and the experiment may not be fully informative. Private information coupled with a partially informative experiment means that in no event will the voter be certain of the incumbent’s reputation, even when she chooses to implement her ideal policy, because by Bayes’ rule, the reputation after a successful policy is necessarily a function of the initial reputation. Further, because $y_2(b_0)$ is no longer linear, we cannot assume that $E(y_2(b_0))$ is sufficient for the voter’s decision. Instead, voters will have to consider $E(y_2(b_0)|y)$. The incumbent is re-elected if $E(y_2(b_0)|y) > y_2(b_{c1})$, or equivalently, if $y^{-1}(E(y_2(b_0)|y)) > b_{c1}$. Hence, the probability of victory for the incumbent is, $F(y^{-1}(E(y_2(b_0)|y)))$, and thereby the expected probability of victory if action is chosen is

$$E(F|b_0, a) = b_0(F(E(y_2(b_+(b_0))|y_1 = 1)) + (1 - b_0)(F(E(y_2(b_-(b_0))|y_1 = -1)).$$

Hence, the incumbent prefers action if and only if:

$$b_0(F(E(y_2(b_+(b_0))|y_1 = 1)) + (1 - b_0)(F(E(y_2(b_-(b_0))|y_1 = -1)) \geq F(E(y_2(b_0))|y = x_1 = s_1))$$

Note that the left hand side is simply $b_0$ times a constant with respect to $b_0$ plus $1 - b_0$ times a lesser constant with respect to $b_0$. Hence, it is strictly increasing in $b_0$, while the right hand side is constant with respect to $b_0$. Therefore, the incumbent will wish to implement her ideal policy if and only if $b_0$ is sufficiently high, and will prefer the status quo otherwise. This implies the existence of a $b_{PI}$ such that $b_0 < b_{PI}$ implies $x_1 = s_1$ and $b_0 > b_{PI}$ implies $x_1 = \theta_1$. 

44
Hence, the inequality can be rewritten as:

\[ b_0(F(\mathbb{E}(y_2(b_+(b_0))) | b_0 \geq b_{PI})) + (1 - b_0)(F(\mathbb{E}(y_2(b_-(b_0))) | b_0 \geq b_{PI})) \geq F(\mathbb{E}(y_2(b_0)) | b_0 \leq b_{PI}) \]

We now show that \( b_{PI} = 0 \) remains a solution. In that event, the expectations on the left hand side become unconditional, and the expectation on the right hand side becomes degenerate.

\[ b_0F(\mathbb{E}(y_2(b_+(b_0)))) + (1 - b_0)F(\mathbb{E}(y_2(b_-(b_0)))) \geq F(y_2(0)) \]

Note that

\[ \mathbb{E}(y_2(b_-(b_0))) = \int_0^1 g(b_0)y_2(b_-(b_0))db_0. \]

Because \( b_-(b_0) < b_0 \) for every \( b_0 \in (0, 1) \), it must be that \( y_2(b_-(b_0)) < y_2(b_0) \), for every \( b_0 \in (0, 1) \). Hence, \( \mathbb{E}(y_2(b_-(b_0))) > \int_0^1 g(b_0)y_2(b_0)db_0 = \mathbb{E}(y_2(b_0)) \), and further, the least possible value for \( \mathbb{E}(y_2(b_0)) \) is \( y_2(0) \).

By contrast,

\[ \mathbb{E}(y_2(b_+(b_0))) = \int_0^1 g(b_0)y_2(b_+(b_0))db_0. \]

Because \( b_+(b_0) > b_0 \) for every \( b_0 \in (0, 1) \), it must be that \( y_2(b_+(b_0)) > y_2(b_0) \), for every \( b_0 \in (0, 1) \). Hence, \( \mathbb{E}(y_2(b_+(b_0))) = \int_0^1 g(b_0)y_2(b_+(b_0))db_0 > \int_0^1 g(b_0)y_2(b_0)db_0 = \mathbb{E}(y_2(b_0)) \), and the least possible value of \( \mathbb{E}(y_2(b_0)) \) is \( y_2(0) \). Hence, both \( F(\mathbb{E}(y_2(b_+(b_0)))) \) and \( F(\mathbb{E}(y_2(b_-(b_0)))) \) are strictly greater than \( F(y_2(0)) \), and so

\[ b_0F(\mathbb{E}(y_2(b_+(b_0)))) + (1 - b_0)F(\mathbb{E}(y_2(b_-(b_0)))) \geq F(y_2(0)) \]

for every \( b_0 \in [0, 1] \), and therefore every type of incumbent prefers to implement her ideal policy. Hence, \( b_{PI} = 0 \) is an equilibrium.
8.3 Mathematical Appendix

Lemma 9. Let \( H : X \subset \mathbb{R} \to \mathbb{R} \) be a continuous and differentiable function.

(i) If there is a unique root \( \bar{x} \) for and \( H'(\bar{x}) > 0 \), then \( H \) must be positive for all \( x > \bar{x} \) and negative for all \( x < \bar{x} \). If \( H'(\bar{x}) < 0 \), then \( H \) must be negative for all \( x > \bar{x} \) and positive for all \( x < \bar{x} \).

(ii) If there are multiple roots, ordered \( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n \), then if \( H'(\bar{x}_i) > 0 \), \( H \) must be positive for all \( x \in (\bar{x}_i, \bar{x}_{i+1}) \). If \( H'(\bar{x}_i) < 0 \), \( H \) must be negative for all \( x \in (\bar{x}_i, \bar{x}_{i+1}) \).

Proof. In case one, \( H'(\bar{x}) > 0 \) implies that for \( x > \bar{x} \) but sufficiently close, \( H(x) > 0 \). Because there is a unique root, there is no \( x > \bar{x} \) for which \( H(x) = 0 \), and thus, by continuity of \( H \), it also cannot be the case that \( H(x) < 0 \) for any \( x > \bar{x} \). Therefore, \( H(x) > 0 \) for all \( x > \bar{x} \). Similarly, \( H'(\bar{x}) > 0 \) implies that for \( x < \bar{x} \) but sufficiently close, \( H(x) < 0 \). Because there is a unique root, there is no \( x < \bar{x} \) for which \( H(x) = 0 \), and thus, by continuity of \( H \), it also cannot be the case that \( H(x) > 0 \) for any \( x < \bar{x} \). Therefore, \( H(x) < 0 \) for all \( x < \bar{x} \). The situation in which \( H'(\bar{x}) < 0 \) is symmetric.

In case two, \( H'(\bar{x}_i) > 0 \) implies that for \( x > \bar{x} \) but sufficiently close, \( H(x) > 0 \). Because there is no root in \((\bar{x}_1, \bar{x}_2)\), it cannot be the case that there exists \( x \in (\bar{x}_1, \bar{x}_2) \) such that \( H(x) = 0 \), and and thus, by continuity of \( H \), it also cannot be the case that \( H(x) < 0 \) for any \( x \in (\bar{x}_1, \bar{x}_2) \). Therefore, \( H(x) > 0 \) for all \( x \in (\bar{x}_1, \bar{x}_2) \). The situation in which \( H'(\bar{x}_i) < 0 \) is symmetric.

References


