Cover Stories

Michael F. Joseph
UCSD
mfjoseph@ucsd.edu

Matt Malis
IE University
matthew.malis@ie.edu

August 22, 2026

Abstract

How do governments maintain plausible deniability for their secret and controversial policies? Existing research focuses on the risk of direct exposure; we highlight the challenges posed by *circumstantial* evidence, and the inferences that audiences draw from the strategic environment. Through a formal model, we uncover a novel tradeoff that leaders face: when the risk of direct exposure of secret policies is low, strategic audiences are especially likely to attribute an unexplained success to secret action. Governments can resolve this attribution problem by using a "cover story"—a public action that the government privately knows to be ineffective, taken alongside a secret action, which provides an alternative explanation for a policy success. We trace this mechanism through detailed examination of the CIA's Operation PBSUCCESS (Guatemala, 1954), with additional evidence from Australian-Timorese treaty negotiations and the Cuban Missile Crisis, and we suggest applications to ongoing debates in American and comparative politics.

Word count: 11,840

In March 1960, the CIA began organizing Cuban exiles to oust Fidel Castro. Eisenhower demanded that the CIA take extraordinary precautions to avoid direct evidence of US involvement (Poznansky, 2020). But as CIA agents were secretly meeting Cuban contacts and building bases in Guatemala and Florida, Eisenhower initiated a public showdown with Castro. In December 1960, Eisenhower announced a complete elimination of Cuba's sugar import quota, justified by Cuba's "deliberate hostility" towards the US and increasing economic integration with the Soviet bloc (Eisenhower, 1960). The next month, the administration formally severed diplomatic ties with Cuba (DoS, 2023)—a wholly symbolic gesture, as the US ambassador had already been recalled and government communications suspended (ADST, 2023, p.53-59). Shortly after, the New York Times began reporting on speculations that the CIA could be training and equipping an invasion force (New York Times, 1961; Brewer, 1961*b*).

Why would Eisenhower choose to attract attention while implementing a deeply controversial policy he wanted to keep secret? Conventional wisdom dictates that he would not. A substantial body of research argues that governments maintain plausible deniability by avoiding *direct* evidence of their secret policies (Spaniel and Poznansky, 2018; Joseph and Poznansky, 2018; Carnegie, 2021), and that overt actions can attract scrutiny that raises the risk of exposure (Carson, 2018; Poznansky, 2020; Colaresi, 2012).

While important, the risk of direct exposure is not all that governments consider. We uncover a countervailing incentive whereby governments, counterintuitively, pursue overt actions to plausibly deny their secret actions. We arrive at our insight through a novel conceptualization of *plausible deniability* (Poznansky, 2022), focusing on audiences' ability to draw *strategic inferences* of government behavior. When an audience observes a change in the world which they knew the government wanted, they do not rely only on direct evidence to determine whether the government undertook a secret policy to induce that change. They also form inferences on the basis of *circumstantial evidence*—their knowledge of the government's interests and capabilities, and of the broader policy context. Even when the government successfully avoids direct evidence of secret policies, it still faces a risk of attribution via strategic inferences.

We argue that governments can offset strategic inferences with a *cover story*—an overt action which provides an alternative explanation of how the government achieved the outcome it wanted, without having secretly resorted to means that the audience disapproves of. Before a secret policy has succeeded, public statements and actions may draw attention and heighten the risk of exposure. But after the policy has succeeded, those same public actions can reduce observers' retrospective suspicion that the outcome was achieved via secret means.

We develop a formal model that studies how cover stories can resolve a common strategic problem studied by scholars of secrecy and international security (Spaniel and Poznansky, 2018; Canfil, 2022; Smith, 2019; Colaresi, 2012; Carnegie, 2021; Yoder and Spaniel, 2022; Kurizaki, 2007; Bils and Smith, 2025). In the model, a government can achieve a policy objective through two means: a public action that an audience directly observes, and a secret action that is only observed probabilistically. The government holds private information regarding the efficacy and cost of each policy lever. The audience finds the secret policies more objectionable, and wants to prevent their use. The government's challenge is to achieve its policy objectives, while maintaining plausible deniability for any actions that the audience disapproves of.

Our analysis shows that the nature of the government's plausible deniability challenge varies as a function of *transparency*—the probability that direct evidence of covert action is exposed. Under high transparency, the pressing concern is direct exposure, and we recover the standard logic that the risk of exposure largely deters governments from taking covert action (Spaniel and Poznansky, 2018; Joseph and Poznansky, 2018). As transparency decreases, however, the government faces a tradeoff: covert action becomes less likely to be directly exposed, but a successful outcome is more likely to be attributed to covert action through circumstantial evidence. Under low transparency, if the audience observes a successful policy outcome despite public inaction, they will infer that covert action was likely taken out of public view, and punish the government just as if direct evidence had been revealed.

To overcome this problem, the government uses performative overt action as a cover story. The cover story functions by exploiting the leader's informational advantage regarding the efficacy of her policy levers, and the audience's inability to perfectly attribute outcomes to actions. From the audience's perspective, the observation of a policy success following public action is consistent with two distinct explanations: the public action was effective, and covert action was unnecessary to cause the policy success; or the public action was ineffective, but taken alongside unobserved covert action which actually caused the outcome. Anticipating this ambiguity, a government engaging in covert action can take public action that it privately knows to be ineffective, mimicking a government whose public action is effective and pursued in good faith. A rational audience maintains some suspicion that covert action was used—but importantly, less than what would arise from a successful outcome without public action. The cover story succeeds if it reduces audience suspicion enough to avoid backlash.

We trace the cover story mechanism through an in-depth case study of Operation PBSUCCESS, Eisenhower's covert intervention to oust Guatemalan President Jacobo Arbenz in 1954. Our analysis highlights public actions by the US which we argue cannot be fully explained by the administration's desire to use all available means to advance their policy objectives. Rather, we propose that these actions are best understood as part of a cover story. We show that key decision-makers expressed concerns over strategic inferences; that these concerns motivated the use of ineffective public action; that administration officials drew attention to their public actions in order to disclaim responsibility for covert action; and that observers at the time found the cover story to be convincing.

Finally, two shorter empirical vignettes illustrate the use of cover stories across diverse political contexts: economic negotiations between Australia and East Timor over oil concessions in the Timor Gap, and the resolution of the Cuban Missile Crisis. This extends the theory's domain across multiple state actors, and to several different forms of policy objectives and secretive policy actions.

This study enriches our understanding of many coercive practices with ambiguous attribution (Baliga, Bueno de Mesquita and Wolitzky, 2020)—including secret proliferation (Debs and Monteiro, 2014), rogue state management (Coe, 2018), cyber conflict (Axelrod and Iliev, 2014), and election meddling (Levin, 2020)—by demonstrating how the use of overt action can complicate attribution of covert action in ways not previously considered. It also contributes to the broader research on political agency and accountability (Ashworth, 2012). This literature has rationalized counterintuitive behaviors such as pandering and "fake leadership" (Canes-Wrone, Herron and Shotts, 2001; Maskin and Tirole, 2004), "showing off" (Judd, 2017), admitting ignorance (Backus and Little, 2020), and adopting extreme ideological stances (Izzo, 2023). We introduce a novel feature to the setup of our model—allowing leaders to use both overt and secret policy levers, in isolation or in combination—which likewise yields novel insights into counterintuitive governing behavior: explaining why leaders implement, and broadly publicize, ineffective policies.

1 Secrecy and Plausible Deniability

We examine a setting where decision-makers desire both a policy objective and approval from an audience (e.g. voters, legislators, foreign allies). The audience is generally accepting of the policy objective (e.g. thwarting communism, dismantling drug cartels), but views some means to achieve it as more controversial than others (e.g. diplomatic pressure vs. assassinations; border security vs. gunwalking). If the government finds it infeasible to achieve the objective via the less controversial means, it may secretly pursue the more controversial means, and attempt to conceal its actions from the audience.

These features characterize many policy contexts. For example, the US public generally wants the government to control immigration at the southern border, but not by treating migrants—especially children—inhumanely. The first Trump administration successfully concealed its family separation policy from the US and foreign publics for several months (Horowitz, 2021). The public wants the government to support scientific advancements, but not at the expense of unethical research practices. Government scientists during the 1950s–1970s thus administered experiments on remote, marginalized communities, to conceal the controversial practices that led to scientific breakthroughs (ACHRE, 1996; Joseph and Poznansky, 2024).

While the policy domain is broad, existing research on secret government action is primarily advanced by scholars of international security, with a particular focus on covert intervention (Spaniel and Poznansky, 2018; Poznansky, 2020; Carnegie, 2021). Thus for concreteness, we characterize the government in our theory as an "Intervener" that seeks to influence political developments in a foreign country, and the secret policies they pursue as a controversial covert operation. The Intervener is accountable to a domestic or foreign audience[1] that deems the covert action to be "unscrupulous", and in conflict with the audience's values. The Intervener worries that an exposed covert action will tarnish its reputation, resulting in some form of punishment from the audience.

Consistent with prior research, we assume that the Intervener is concerned with both achieving a successful policy outcome, and maintaining plausible deniability for actions that an audience finds objectionable. We depart in our assumptions about what plausible deniability requires. Existing

[1]Our focus differs from that of Bloch and McManus (2024), who are primarily interested in audiences within the target state of a nominally covert attack, and how the perpetrator's noncredible denial dampens those audiences' desire to retaliate.

theories focus on *direct* evidence as the determinant of whether plausible deniability is maintained. In a comprehensive review, Poznansky (2022, 523-524) identifies three "threats to plausible deniability" at the state level: leaks, rival intelligence, and information and communication technology—all variants of direct evidence. In the two game-theoretic analyses most similar to ours, Spaniel and Poznansky (2018) and Canfil (2022) both assume that a cost is automatically imposed on the administration when covert action is revealed, but do not allow for reputational costs arising endogenously from audience inference or speculation.

We argue that audiences are clever, creating a previously unexplored barrier for sustaining plausible deniability. Specifically, audiences draw inferences from the strategic context. This includes their knowledge of the Intervener's preferred policy outcome, its capabilities to achieve that outcome through unobservable actions, and the likelihood that the outcome would occur in the absence of intervention. We argue that Interveners can use a *cover story* to offset these strategic inferences and maintain plausible deniability.

2 Characterizing a Cover Story

We start with a broad conceptual definition:[2]

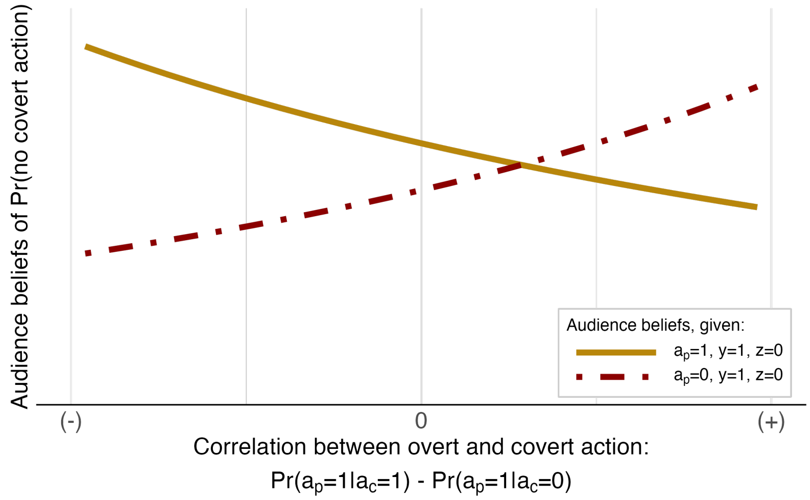
Definition 1 *A* ***cover story*** *is a publicly observable government action which the government can later reference to plausibly disclaim a secret action.*

We now describe a simple model of audience learning about government policymaking, to identify the core conditions under which a cover story succeeds. The model takes the government's behavior as given, and asks what the audience can infer about unobserved behavior: given a policy success, does the audience believe that the government took an unobserved covert action to achieve it? If the observation of an overt action reduces the audience's suspicion that covert action was used, we say that the overt action served as a successful cover story.

The model's essential features are as follows (Appendix 7.1 presents technical details). The government takes either public action ($a_p \in \{0, 1\}$), covert action ($a_c \in \{0, 1\}$), both, or neither.

[2] We note that common usage of the term "cover story" may emphasize a narrative or rhetorical component; our definition, in contrast, highlights the conditions under which such a narrative would be convincing to an audience.

Figure 1: Audience Beliefs given Circumstantial Evidence



Note: Figure is constructed with additive effects of public and covert action. See Appendix 7.1 for details, and for other examples in which the actions are complements or substitutes.

When covert action is taken, it is directly exposed ($z = 1$) with some known probability. Both actions contribute probabilistically to a policy success (denoted $y = 1$) or failure ($y = 0$).

The audience observes a history $h = (a_p, y, z) \in \{0, 1\}^3$, and forms a posterior belief of the probability that the government did *not* use covert action, $\eta^{(a_p,y,z)} = Pr(a_c = 0|(a_p, y, z))$. When covert action is directly revealed, there is no uncertainty in the audience's belief ($\eta^{(a_p,y,z=1)} = 0$); but when the audience does not observe direct evidence of covert action ($z = 0$), their beliefs depend on *circumstantial evidence*—their knowledge of the probabilities of each action profile (given in (5)), and of the policy success function (given in (6)—both equations appearing in Appendix 7.1).

With this setup, we obtain the following general result:

Result 1 *If public and covert action are not too complementary, and their use is not too positively correlated, then overt actions increase the probability of policy success and function as cover stories.*

The two curves in Figure 1 visualize the audience's posterior beliefs that covert action was *not* taken, $\eta^{(a_p,y=1,z=0)}$, given that they observed a policy success ($y = 1$) and no direct evidence of covert action ($z = 0$), under two different scenarios: when public action was taken ($a_p = 1$, solid yellow line), and when public action was not taken ($a_p = 0$, dot-dashed red line).

Following Definition 1, the public action functions as a cover story when the yellow curve is above the red curve—meaning that the observed public action reduces the audience's suspicion that an unobserved covert action was also used. Figure 1 shows that this occurs whenever covert and public action are negatively correlated, or even slightly positively correlated. Under these conditions, when the audience observes a policy success, they infer that *some* government action likely contributed to it. Absent an observed public action, the audience is more likely to attribute the unexplained policy success to an unobserved covert action.

Result 1 shows how cover stories can emerge under general conditions. The analysis so far has been agnostic as to *why* the government takes these actions, or what the government's objective function is. The next section incorporates strategic government behavior, and investigates whether and when cover stories can still succeed in influencing audience beliefs.

3 A Strategic Model of Covert Action and Cover Stories

3.1 Model Setup

We study a strategic interaction between a leader L of an Intervener state, and an audience A who can hold the leader accountable. Figure 2 presents the timing of the game. We discuss each step in turn.

Leader types. To incorporate the reputational considerations discussed above, we assume the leader has a privately known type, $\theta \in \{0, 1\}$, which determines how intrinsically costly the leader finds covert action to be: $\theta = 1$ denotes a "scrupulous" type, and $\theta = 0$ denotes an "unscrupulous" type. In the first step of the game, this type is drawn by nature and observed privately by the leader; the audience holds a prior belief that $Pr(\theta = 1) = \pi \in (\frac{1}{2}, 1)$.

Policy options. The leader has two policy levers: a public (overt) action $a_p \in \{0, 1\}$, which is observed directly by the audience; and a covert (secret) action $a_c \in \{0, 1\}$, which is only observed probabilistically. The leader can enact either one, both, or neither of these policy levers. Referring back to our opening anecdote for concreteness, a_p can represent the Eisenhower administration's imposition of economic pressure on Cuba through oil embargoes and slashing sugar quotas, while a_c can represent CIA efforts to assassinate Castro or oust him via support of Cuban exiles.

Before L decides which policies to authorize, in step 2, L receives private information about

1. The leader L's type $\theta \in \{0, 1\}$ is realized by Nature and observed privately by L.
2. The state variable $\omega \in \{0, 1\}$, and the cost variable $k_c \sim F^{\theta}$, are realized by Nature and observed privately by L.
3. The leader chooses whether to take public action $a_p \in \{0, 1\}$ (which the audience, A, observes), and covert action $a_c \in \{0, 1\}$ (which A does not observe directly).
4. The policy outcome $y \in \{0, 1\}$ is realized, according to the probabilities given in (1).
5. The covert revelation $z \in \{0, 1\}$ is realized, according to the probabilities given in (2).
6. A observes $(a_p, y, z) \in \{0, 1\}^3$, and chooses whether to punish or reward L, $r \in \{0, 1\}$.

Figure 2: Game Sequence

each of the policy options. First, L learns about the effectiveness of public action, denoted by the state variable $\omega \in \{0, 1\}$: the leader observes ω privately, and the audience holds a prior belief that $Pr(\omega = 1) = \tau \in (0, 1)$. Second, the leader learns the cost k_c of covert action, which is drawn from a type-specific distribution $F^{\theta}(x) = Pr(k_c \leq x|\theta)$. We assume that F^0 is continuously differentiable with positive density on a wide interval $(\underline{k_c}, \overline{k_c})$ (as specified in Assumption 1; see Appendix 7.3), and that F^1 has a lower bound of $(1 + \beta)$; this restriction on F^1 means that covert action is always prohibitively costly for the scrupulous leader.

Substantively, step 2 represents the leader's advisers briefing her about the available policies, and the advisers' best assessments of the benefits and drawbacks of each. Returning to Eisenhower's Castro policy, now-declassified documents indicate that the CIA assessed that the Soviets would effectively mitigate the impacts of US economic pressure tactics (FRUS, 1960). As such, NIE 85–2–60 argued that "Castro will almost certainly remain in power through 1960", despite the overt policies being pursued (CIA, 1960). Notably, these intelligence assessments were *private*, and foreign and domestic audiences faced uncertainty as to whether such policies could actually contribute to Castro's downfall. Indeed, one Cuban exile leader—while dismissing the possibility of military intervention—stated in January 1961 that Castro was likely to fall within three months "in view of the economic paralysis and growing discontent" among the Cuban people (Brewer, 1961*a*).

In step 3, the leader chooses which of the policy options to pursue, $a = (a_p, a_c) \in \{0,1\}^2$.

Policy outcomes. In step 4, the policy outcome $y \in \{0,1\}$ is realized, where $y = 1$ denotes success, and $y = 0$ denotes failure. The policy outcome is a probabilistic function of the leader's action a, and the effectiveness ω of the overt action. The policy production function is

$$Pr(y=1|a,\omega) = \begin{cases} \alpha_{pc}^{\omega}, & a_p = 1 \,\&\, a_c = 1 \\ \alpha_p^{\omega}, & a_p = 1 \,\&\, a_c = 0 \\ \alpha_c, & a_p = 0 \,\&\, a_c = 1 \\ \alpha_0, & a_p = 0 \,\&\, a_c = 0 \end{cases}, \tag{1}$$

where α_0 denotes the probability of success due to exogenous factors. We assume $0 < \alpha_0 \leq \alpha_p^0 < \alpha_p^1 < 1$: public action (weakly) increases the probability of success in either state ($\alpha_p^{\omega} \geq \alpha_0$); it is more effective when $\omega = 1$ than when $\omega = 0$ ($\alpha_p^0 < \alpha_p^1$); but it never guarantees success ($\alpha_p^{\omega} < 1$). Likewise, we assume $\alpha_0 < \alpha_c < 1$. When covert and public action are taken together, the probability of success is[3]

$$Pr(y=1|a_p = a_c = 1, \omega) = \alpha_{pc}^{\omega} = \alpha_p^{\omega} + (1-\alpha_p^{\omega})\alpha_c$$

We note that this specification is compatible with covert and public actions being either complements or substitutes in the policy production function.[4]

Covert revelation. The leader initially takes covert action in secret, but accepts some irreducible risk of direct exposure. In step 5, $z \in \{0,1\}$ denotes whether covert action is exposed, with

$$Pr(z=1|a) = a_c\lambda \tag{2}$$

Whenever L refrains from covert action, A observes $z = 0$; but if L does take covert action, A observes $z = 1$ with probability $\lambda \in (0,1)$. We refer to λ as the level of *transparency* in the policymaking environment.

[3]Similar to Spaniel and Poznansky (2018), we consider the probability of failure as the joint probability of both covert and public action failing independently: $Pr(y=0|a_p = a_c = 1, \omega) = (1-\alpha_p^{\omega})(1-\alpha_c) = 1 - \alpha_{pc}^{\omega}$.

[4]The two policy levers are complements if $\alpha_{pc}^{\omega} - \alpha_p^{\omega} > \alpha_c - \alpha_0$ (equivalently, $\alpha_0 > \alpha_p^{\omega}\alpha_c$).

Audience punishment/reward. In step 6, the audience chooses whether to punish ($r = 0$) or reward ($r = 1$) the leader. The audience receives a payoff of 1 for rewarding a scrupulous leader, or for punishing an unscrupulous leader, and 0 otherwise:

$$U_A = \mathbb{1}[r = \theta] \tag{3}$$

The substantive interpretation of this action depends on the audience. For a domestic voter, r can represent the choice to support the incumbent leader against her electoral challenger (Myrick, 2020). For a legislature, it can represent the choice to punish the executive by withholding funding of its policy priorities, curtailing its statutory authority, or conducting embarrassing public investigations (Colaresi, 2012; Spaniel and Poznansky, 2018; Smith, 2019). For foreign audiences, r can represent the choice over whether to cooperate with L on future foreign policy initiatives, or to withdraw from L's bloc or alliance system more broadly (Wallace, 2024; Poznansky, 2025). In each case, the audience prefers to "reward" the leader if and only if the leader is scrupulous, but faces uncertainty about the leader's true type. We assume $\pi > \frac{1}{2}$, meaning the leader enjoys a "presumption of innocence": the audience is disinclined to punish the leader based on their prior beliefs of her type, but may be swayed toward punishment on the basis of direct or circumstantial evidence.

Leader's payoff. The leader's payoff is

$$U_L = y - a_c k_c - a_p k_p + r\beta \tag{4}$$

Both leader types enjoy a benefit normalized to one for a successful policy outcome (and zero for failure); they receive a reputational benefit of $\beta \in (0, 1)$ when rewarded by the audience (with the "penalty" of punishment normalized to zero); and they pay direct costs k_p and k_c for taking public action and covert action, respectively. As stated above, the factor that distinguishes the two types of leader is the distribution from which their covert action cost k_c is drawn. Scrupulous leaders always find covert action to be prohibitively costly. Unscrupulous leaders are less intrinsically opposed to taking covert action; whether or not they do so depends on the realization of k_c, and the incentive scheme created by the audience's endogenous punishment/reward strategy.

3.2 Analysis

When a leader's most effective policy option is a controversial covert action, how does she both achieve a successful policy outcome and maintain plausible deniability for the actions she took to achieve it? Our analysis reveals that the leader's optimal strategy depends on transparency, with novel behavior emerging when transparency is low.

We begin by characterizing the novel equilibrium behavior that our analysis will reveal.

Remark 1 *If the direct costs of public action outweigh the direct policy benefits—that is, in the unfavorable state, $\omega = 0$—a cover story motivation (per Definition 1) is necessary for the leader to take public action.*

Our model is structured so that the leader has two possible incentives for public action: increasing the probability of a policy success, and maintaining a favorable reputation with the audience. When $\omega = 0$, the direct costs of public action outweigh the direct benefits,[5]

$$k_p > E[y|a_p = 1, a_c, \omega = 0] - E[y|a_p = 0, a_c, \omega = 0], \quad a_c = 0, 1,$$

meaning that public action cannot be justified by policy incentives alone. Thus when the leader takes *ineffective* public action, we can be certain that she would not have done so in the absence of reputational incentives.[6]

Definition 2 (CSE) *The leader employs a "pure" cover story by taking public action when the direct costs outweigh the direct policy benefits, while also taking covert action.*[7] *A* ***cover story equilibrium (CSE)*** *is an equilibrium in which a pure cover story is played with positive probability.*

For the remainder of the game-theoretic analysis, our use of the term "cover story" will refer to the "pure" case described in Definition 2. A cover story (as defined in Definition 1) can arise under more general conditions, per Result 1; but the narrower focus of Definition 2 highlights the theoretically hard case for cover stories to emerge in equilibrium, and presents clear, ex-ante

[5]Assumption 1 imposes the restriction that $\alpha_p^0 < k_p < \min\{\alpha_p^1 - \alpha_0, \alpha_p^1(1 - \alpha_c)\}$.

[6]More precisely, we may think of public action as being *cost-ineffective* when $\omega = 0$, as it can still improve the probability of success ($\alpha_p^0 \geq \alpha_0$).

[7]Formally, $\Pr(\text{cover story}) = \Pr(a_p = 1|\omega = 0, a_c = 1)$. Note that $Pr(\omega = 0, a_c = 1)$ is always positive (see Remark 7 in Appendix 7.4.1).

empirical conditions where we are theoretically confident that a cover story is necessary to explain a leader's behavior.

Central to the equilibrium logic is the audience's posterior belief that the leader is scrupulous. This belief depends on the risk of direct exposure of covert action, represented by the transparency parameter λ: the level of transparency not only determines whether the audience observes evidence of covert action directly, but also affects the inferences that the audience can draw in the absence of any direct evidence. We can thus characterize the model's equilibrium as a function of thresholds in this parameter.[8]

Proposition 1 *There exist thresholds $\bar{\lambda}$ and $\bar{\bar{\lambda}}$, with $\bar{\lambda} \leq \bar{\bar{\lambda}}$, such that:*[9]

- *If transparency is high ($\lambda \geq \bar{\bar{\lambda}}$), the audience does not punish the leader in the absence of direct exposure of covert action.*
- *If transparency is in an intermediate range ($\bar{\lambda} < \lambda < \bar{\bar{\lambda}}$), the leader does not use a cover story, and the audience may punish the leader for an "unexplained success".*
- *If transparency is low ($\lambda < \bar{\lambda}$), the leader uses a cover story with positive probability.*[10]

The following sections discuss the intuition behind the equilibrium behavior across these regions. Proofs for all formal results are presented in the appendix.

3.2.1 Trust and restraint under high transparency

From A's utility function (3), we see that the audience's equilibrium strategy depends on their posterior belief of the leader's quality. Given the observed history of the game $h = (a_p, y, z) \in \{0, 1\}^3$, the audience forms a belief $\mu^h = Pr(\theta = 1|h)$. A prefers to reward the leader ($r = 1$) if $\mu^h > \frac{1}{2}$, and punish ($r = 0$) if $\mu^h < \frac{1}{2}$. Recall that the scrupulous leader never takes covert action. Thus, upon directly observing covert action ($z = 1$), the audience draws the most negative possible inference of the leader's scrupulousness ($\mu^{a_p, y, z=1} = 0$), and punishes her accordingly.

[8] Our analysis imposes parameter restrictions (Assumption 1) and off-path beliefs (Assumption 2), stated in Appendix 7.3. For expositional clarity, the main text focuses on a particular class of equilibrium, which we label "interior Responsive Equilibria"; see Appendix 7.4.2.

[9] Note that $\bar{\lambda} = \bar{\bar{\lambda}}$ only if $\bar{\lambda} = \bar{\bar{\lambda}} = 0$; see Proposition 4 in the Appendix.

[10] In the knife-edge case of $\lambda = \bar{\lambda}$, the no-cover equilibrium exists, and may coexist with a CSE.

Absent direct exposure ($z = 0$), the audience's inferential challenge is more complex. To build intuition, first suppose that transparency is high ($\lambda \geq \bar{\bar{\lambda}}$). In this case, the leader knows that covert action is very likely to be exposed, which incentivizes her to take covert action only rarely (when k_c is very low). Understanding this, the audience in turn infers that if covert action was not exposed, it was unlikely to have been taken; the absence of evidence of covert action provides compelling (though not definitive) evidence of the absence of covert action. Thus in any $z = 0$ history, the audience maintains a posterior belief of $\mu^{a_p,y,z=0} > \frac{1}{2}$, and does not punish the leader. These beliefs (specifically in the $y = 1$ histories), along with the low probability of covert action, are visualized on the righthand side of Figure 3.

3.2.2 The problem of unexplained success

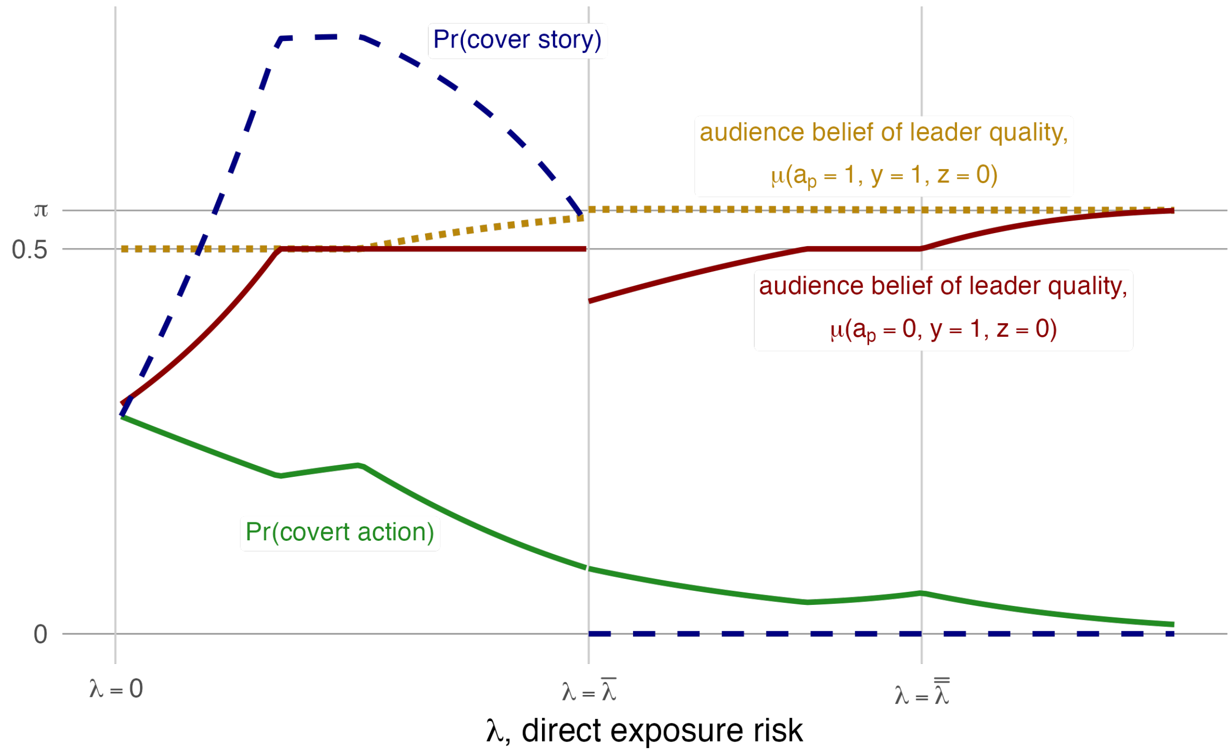
As λ decreases towards $\bar{\bar{\lambda}}$, two forces push against the audience's favorable posteriors. First, the audience knows that the leader is taking covert action with increasing frequency; and second, the audience knows that the absence of direct exposure is an increasingly unreliable signal. Once λ falls into the intermediate range ($\lambda \in \left(\bar{\lambda}, \bar{\bar{\lambda}}\right)$), these forces combine to make the audience no longer willing to give the leader the "benefit of the doubt" in all instances of ambiguous attribution.

The history that makes the audience most suspicious is that of an "unexplained success": a successful outcome with no public action, and no direct evidence of covert action (formally, $h = (a_p = 0, y = 1, z = 0)$). Upon observing an unexplained success, the audience infers that one of two things must have happened: either the leader took no action, and the success arose due to random luck (which occurs with probability α_0); or the leader took covert action, but no direct evidence came to light. The relative weight that the audience assigns to each possibility depends on the level of transparency.

Under intermediate transparency, the audience's suspicion leads them to (sometimes) punish the leader following an unexplained success. This threat of punishment has some disciplining effect on the leader's behavior, visualized in Figure 3 as a slight decrease in $Pr(a_c)$ for λ just below $\bar{\bar{\lambda}}$. But as λ continues to decrease, the audience's belief $\mu^{0,1,0}$ becomes increasingly unfavorable, falling below the threshold of $\frac{1}{2}$, and they fully punish the leader purely on the basis of circumstantial evidence.

At the same time, in the intermediate range of transparency, the audience's belief is more

Figure 3: Leader Strategy and Audience Beliefs



Note: Figure depicts leader strategies and audience beliefs in an equilibrium characterized by Proposition 1. Pr(covert action) is conditional on unscrupulous type, and Pr(cover story) is conditional on covert action and $\omega = 0$. π is the prior probability the leader is scrupulous; $\mu(a_p, y, z)$ denotes the audience's posterior belief. Beliefs $\mu(a_p, y = 0, z = 0)$ sit above 0.5 for the full range of λ in this equilibrium, and are omitted for visual clarity. Figure is constructed with $\alpha_0 = 0.05, \alpha_p^0 = 0.1, \alpha_c = \alpha_p^1 = 0.5, k_p = 0.2, \beta = 0.5, \tau = 0.5, \pi = 0.55, F^0 = N(\text{mean} = 0.3, \text{sd} = 0.15)$.

favorable after seeing the leader take public action. This is visualized in the gold dotted line in the figure remaining above $\frac{1}{2}$ across the range of $\lambda > \bar{\lambda}$. As a result, the audience does not punish the leader when they observe public action under intermediate (or high) transparency.

3.2.3 The value of the cover story

The two aforementioned features of the audience's strategy—punishing unexplained success, but rewarding public action—provide the rationale for the leader's use of a cover story. Consider the unscrupulous leader's evaluation of her policy options when she (privately) expects public action to be ineffective ($\omega = 0$), but a relatively low-cost covert action is available (k_c low). She could pursue the covert action alone, accepting that if it succeeds, she will face severe backlash from the audience whether or not direct evidence comes to light. Alternatively, in addition to pursuing the covert action, she could also take the ineffective public action, claiming that she actually expects it to be effective. If the successful outcome is achieved, and direct evidence of covert action remains unexposed, she can point to the public action as the cause of the policy success and hope that her audience is willing to accept that explanation. This is the strategic logic of the cover story.

Remark 2 *The successful use of the cover story (per Definition 2) depends on the audience's uncertainty over the effectiveness of public action.*

Informational asymmetry between the leader and the audience is critical to the functioning of our "pure" cover story mechanism. Suppose that the audience did know the value of ω, and thus knew whether a leader's public action was effective or ineffective. Within the CSE, upon observing the leader take ineffective public action, the audience would know that the public action was serving as a cover story, and would immediately infer that the leader was also taking covert action. What enables the cover story to work to the leader's advantage is the audience's residual uncertainty over ω. After observing public action, the audience assigns some probability to it having been effective and pursued in good faith, and some probability to it having been deployed as a cover story.[11]

The following corollary formally outlines the conditions under which a cover story is a worthwhile gambit for the leader:

Corollary 1 (CSE comparative statics) *The leader uses a cover story with positive probability when transparency is low, $\lambda < \bar{\lambda}$. The threshold $\bar{\lambda}$ is:*

- *increasing in the effectiveness of covert action, α_c;*
- *decreasing in the direct cost of public action, k_p;*
- *and, if α_0 is low, $\bar{\lambda}$ is increasing in the leader's value for audience approval, β.*

The leader's use of a cover story entails a tradeoff between the direct cost of public action (k_p), and the reputational benefits bestowed by the audience (β). The cover story only serves to *improve* the leader's reputational payoff (relative to taking covert action on its own) if (i) covert action is not directly exposed (which occurs with probability $1-\lambda$), and (ii) the policy succeeds (probability $\alpha_p^0 + \alpha_c(1-\alpha_p^0)$ with the public action, α_c without), since the audience has little reason to suspect

[11] See Corollaries 3 and 4 in Appendix 7.4 for further discussion.

covert action given an unsuccessful outcome. Thus the net benefit of a cover story is increasing in α_c and in β, and decreasing in k_p and in λ.

Technically, our model setup assumes that the audience observes a_p, y, and z simultaneously. In reality, these observations are likely sequenced: the audience first observes the leader's action, then later observes the policy outcome and (possibly) direct evidence of covert action. Considering the audience's interim beliefs—after observing the leader's action but before observing the outcome—can provide further substantive insight into what the cover story accomplishes for the leader.

Corollary 2 (Cover Stories and Scrutiny) *In a CSE, the audience's interim beliefs of the leader's scrupulousness after observing public action (but before observing the outcome) are strictly less favorable than their interim beliefs after observing no public action.*

Either the state is favorable ($\omega = 1$) and both leader types would take public action, so the audience learns nothing about the leader's type from observing public action; or the state is unfavorable ($\omega = 0$), and only the unscrupulous type would take public action. As long as there is positive probability of the latter, the audience updates negatively about the leader's type when they observe public action.

Substantively, this result can explain the intuition—as suggested in the anecdote of Eisenhower's Castro policy—that public action itself can "draw attention" to an issue, and make rational audiences more suspicious that covert action is afoot. Our analysis demonstrates that, despite raising short-term suspicion, cover stories provide leaders a long-term explanation for how events turned in their favor without their having resorted to unscrupulous covert action.

3.3 Empirical Implications for Covert Action Research

One implication of our theory highlights the strategic challenges faced by policymakers concerned with plausible deniability:

Implication 1. In any case of covert action, mission planners within Intervener states should not only be concerned with operational security and the risk of direct exposure; they should also consider how they are perceived by a skeptical audience—even in the best-case scenario that the operation succeeds and no direct evidence comes to light—and how they might be able to allay the audience's suspicion of their involvement. Further, where mission planners are *least* concerned with the risk of direct exposure, they should be *most* concerned about being blamed by the audience for an unexplained success.

We predict that when Interveners use covert action under conditions of low transparency, they will deploy a cover story. In such cases, we expect:

Implication 2. The Intervener will use overt action as a cover story for their covert intervention. The overt action should be less objectionable to the audience than the secret policies that the leader hopes to cover up. It should not be too intrinsically costly (k_p low), and it should be plausible, from the audience's perspective, that the action could significantly contribute to the likelihood of policy success ($\alpha_p^1 > \alpha_0$).

Implication 3. The Intervener should make an effort to connect the favorable policy outcome to the public action in the mind of the audience.

Implication 4. The audience should be convinced by the cover story, to a sufficient degree that they are willing to refrain from punishing the leader. Specifically, they should infer that covert action was unlikely to have been used; and this inference should be based partly on the absence of direct evidence, and partly on the observation of public actions which ostensibly contributed to the policy success.

3.4 Alternative Theories of Joint Covert and Overt Actions

We briefly discuss two potential alternative explanations for jointly deployed secret and overt actions.

In a *policy buildup*, a leader pursues different policy options sequentially. She first authorizes the less controversial overt policies; if they fail to achieve the desired outcome, she then ratchets up, eventually authorizing covert action. A key distinction from a cover story is that the leader originally hopes to achieve the outcome through overt actions alone. Covert actions are only pursued once the public actions have been attempted and failed. Thus, in any case where covert action preceded public action, or both were authorized simultaneously, we can be confident in ruling out a policy buildup.

Under *comprehensive pressure*, leaders implement overt and covert policies simultaneously, believing that each action has direct policy benefits which outweigh the associated costs. Leaders may view both covert and public actions as substantially increasing the probability of succeeding in some primary policy objective (for instance, removing a foreign regime from power); or they may view overt actions as useful for advancing some secondary objective, in the event that the primary objective fails (for instance, weakening or isolating a target regime in case it remains in power).

Three implications help to differentiate comprehensive pressure from a cover story. First, comprehensive pressure cannot explain the use of public actions that policymakers have judged as too ineffective to justify the costs. Second, it cannot explain overt actions that draw unnecessary attention without enhancing the chance of success. Third, cover stories and comprehensive pressure may imply divergent impacts of the public action on audience beliefs. If the variants of comprehensive pressure described above were the primary motivation for overt action, we would expect covert and public action to be positively correlated. Following our analysis in Section 2, this would imply that the audience should draw an unfavorable inference regarding the leader's conduct when they observe public action and a policy success.

We evaluate these alternatives against our cover story explanation in the case study that follows.

4 Operation PBSUCCESS

The 1950 presidential election marked the first time in Guatemala's history that power was peacefully transferred from one democratically elected leader to another. From an institutional perspective, the 1950 election suggested that democracy was working (Fraser, 2005, 487). But it was not working for the United States. US policymakers grew alarmed watching newly elected President Jacobo Arbenz appoint communists to government positions (Immerman, 1982, 108) and implement land reforms (Schlesinger and Kinzer, 1982, 53). In his memoirs, Eisenhower worried that "Communism was striving to establish its first beachhead in the Americas by gaining control of Guatemala."[12]

[12]Quoted in Schmitz (1999, 179)

In August 1953, Eisenhower authorized the covert CIA operation PBSUCCESS. The first phase involved establishing bases in neighboring countries, which would be used to train and arm 480 Guatemalans to overthrow the Arbenz government. The CIA also groomed a staunch anti-communist and former coup-plotter, Castillo Armas, to lead the rebellion. But the genius of the plan lay in the psychological operations (Cullather, 2006). Because the CIA was skeptical that a small paramilitary force alone could overthrow the government, they also developed offensive psychological operations to convince loyalists that defending Arbenz was futile and would lead to reprisals. Such operations also included a media blitz across Latin America, and threats and bribes of Guatemalan politicians to have them recognize the coup plotters as the rightful government (Schlesinger and Kinzer, 1982, 114).

PBSUCCESS is widely seen as a successful covert action. Arbenz resigned on 27 June 1954 in the face of military incursions, and the CIA avoided direct evidence of their involvement. Broadly speaking, the Eisenhower administration retained enough plausible deniability to avoid backlash from foreign or domestic audiences.

Following best practices in case-evaluation of formal models, we conduct a within-unit analysis (Gerring, 2004), mapping several interconnected beliefs, expectations and choices from our theory to the empirical record of the case (see Bates, Greif, Levi, Rosenthal and Weingast, 1998; Lorentzen, Fravel and Paine, 2017; Levi and Weingast, 2022). Following Goemans and Spaniel (2016), Joseph, Poznansky and Spaniel (2022) and others, we examine the historiography and primary source evidence of the Eisenhower administration's decision-making processes, paying particular attention to the choice nodes that we model. We develop case-specific hypotheses regarding the decisions and beliefs that our theory predicts at each moment, and we evaluate them against leading alternatives.

4.1 Case Selection, and Calibrating the Parameters

Following Bates et al. (1998), we identified Eisenhower's Guatemala intervention as a case fitting the core strategic tension characterized by our model. Eisenhower believed that stopping the spread of communism required that the US maintain a reputation for upholding the principles of sovereignty, self-determination and democracy (Rabe, 1988, 166). Yet there remained uncertainty over the US's commitment to these values—reflecting uncertainty over the leader's "type." Eisenhower understood that using military force to overturn a democratically elected government in

Guatemala would "stigmatize our international reputation" (CIA, 1954*b*), and that the exposure of this effort would induce damaging reactions from the US public and international audiences, particularly in Latin America (Schmitz (1999), Jeffreys-Jones (2022, ch. 4), Poznansky (2019)).

The case's initial conditions fit the cover story equilibrium per Corollary 1. The minimal communications technology in Guatemala, along with the CIA's relatively advanced capabilities in ensuring tight operational control (Cullather, 2006, 7), imply transparency was low (λ). The recent success of a similar playbook to oust Mossadeq in Iran undetected gave the Administration confidence that covert intervention was their best chance of success (higher α_c). The preceding discussion, along with Eisenhower's domestic political interests (with midterms that year, and reelection soon after), imply high reputational concerns (high β).

We found theoretically that a cover story was a necessary motivation for overt action if the leader privately assessed that the costs of overt actions alone outweighed their direct policy benefits. At multiple points before PBSUCCESS, the US government considered and ruled out various overt actions against Guatemala. Truman considered economic sanctions and diplomatic isolation in 1949–1950, but "it was decided at this time that such drastic measures were not yet advisable" because they would not create meaningful local pressure for the government to change policy, and could even galvanize Guatemalan support for Arbenz (Immerman, 1982, 218). State Department officials considered and again discarded similar overt measures in 1952, believing that the Soviets were likely meddling covertly, and "it was necessary to fight fire with fire" (Immerman, 1982, 239): some form of US-backed military intervention was necessary to oust Arbenz or alter his policy. The US Ambassador to Guatemala, John Peurifoy, advised in December 1953 that the application of economic sanctions—a seemingly plausible overt means of undermining Arbenz's hold on power—would likely backfire, as it might "damage irreparably the propertied class and prevent it from retarding the advance of Communism" (FRUS, 1953*b*). In short, even though Eisenhower was ultimately willing to authorize overt policies and PBSUCCESS concurrently, overt options were found not viable on their own (Cullather, 2006, 35).

One critical factor shaping these assessments was the absence of local actors willing and able to capitalize on the opportunities that overt US actions could create. Cullather (2006, 34) writes that the "'only organized element in Guatemala capable of decisively altering the political situation,' the Army, showed no inclination toward revolutionary action" (see also Rabe, 1988, 55). Even Castillo

Armas, the CIA's hand-selected leader of the invasion force later in 1954, "showed little inclination to launch his revolution without Agency support" (Cullather, 2006, 33).

4.2 The Puzzle of Overt Action

From the outset of planning Operation PBSUCCESS, plausible deniability was viewed as essential to the mission's success (Immerman, 1982, p133). "Covert accomplishment of the objectives of PBSUCCESS", according to a planning document in late 1953, was to be "defined as meaning accomplishment with plausible denial of United States or CIA participation" (FRUS, 1953*a*) after the operation was concluded. Administration officials repeatedly warned mission planners: "don't get caught" (FRUS, 1954). But with officers stationed across Latin America to train and supply the coup plotters—even opening an operation center inside Guatemala in December 1953 (Cullather, 2006, App. A)—there always remained a risk of direct exposure, particularly after the active phase of PBSUCCESS was given the "full green light" in April 1954.

Given the intense focus on secrecy, we might expect the administration to divert public attention away from Guatemala, so as to minimize the risk of exposure. Yet we observe the opposite. US diplomatic activity leading up to the coup indicated intense interest in political developments within Guatemala. In early 1954, Ambassador Peurifoy and others made inflammatory statements that the US would not tolerate a communist country between Florida and the Panama Canal. In March, at the Caracas Conference of the OAS, US representatives forced an anticommunist resolution designed to isolate Guatemala first on the meeting's agenda (Immerman, 1982). Dulles delivered an impassioned speech at the conference attacking the Guatemalan government. This was at the direction of Eisenhower, who instructed his diplomats that "By every proper and effective means we should demonstrate to the courageous elements within Guatemala who are trying to purge their government of its communist elements that they have the sympathy and support of... the US." By "proper", Eisenhower meant public and short of calling for military intervention (Bowen, 1983).

During the military phase of PBSUCCESS, when the CIA was most exposed, the Administration ramped up their overt policies. On May 15, a freighter carrying weapons that Arbenz had purchased from Czechoslovakia landed in Guatemala (Immerman (1982, 155); Schlesinger and Kinzer (1982, 147)). Arbenz had hoped to keep the shipment secret, but the US discovered it the next day (Cullather, 2006, 80). Rather than minimize the episode, Eisenhower expressed public outrage.

He invoked the Monroe Doctrine and imposed a naval blockade to prevent future arms shipments into Guatemala (Cullather, 2006, 79). In fact, from the US perspective, the Czechoslovakian arms shipment was serendipitous: before discovering the shipment, the CIA had planned to fabricate a Soviet arms cache, under operation WASHTUB (Cullather, 2006, 101), which the US would then "discover" and exploit publicly.

Why would Eisenhower shine a light on US concerns over Guatemala when covert operations were underway? The conventional explanation, to the extent one exists, is comprehensive pressure: Eisenhower authorized all available policies, both overt and covert, to maximize the chance that Arbenz would step down (e.g. Cullather, 2006, p59). But given Eisenhower's focus on maintaining secrecy, this explanation cannot account for *how* and *where* the Administration publicized overt actions as PBSUCCESS was unfolding. For example, while PBSUCCESS relied partly on broadcasting anti-Arbenz messages across Guatemala, mission success did not require that these messages be voiced by American foreign policy elites. In fact, there was concern that "hard hitting speeches against Guatemala by personages in the United States Government could be counter-productive and would particularly alienate those non-Communists whose actions are influenced by nationalist emotions" (CIA, 1954*f*). Contemporary reporting in the New York Times validated this concern, observing that "the outcry of the United States against the arms shipment. . . had produced a solidarity of Guatemalan opinion behind the government that had been surprising even to government leaders themselves!"[13]

Mission success also did not depend on US domestic support. Nonetheless, DCI Dulles deliberately exaggerated the scope of the weapons shipment to prompt Congressional statements and press coverage within the United States (Cullather, 2006, p59). These tactics carried significant risks. Assistant Secretary of State Cabot had previously warned that if US "public opinion should become too aroused and excited, there might be embarrassing demands for [overt] action... [that were] altogether infeasible" (CIA, 1953). Indeed, the increased domestic interest in Guatemala prompted the chairman of an anti-communist congressional committee (who was not read in on PBSUCCESS) to travel to Guatemala a week before the invasion was launched. Dulles recognized this tension: he expressed "special concern. . . that in their eagerness to expose international Communism in Central America, the [congressmen's] effort would 'run afoul of or cross wires with' CIA

[13]Quoted in Taylor (1956, 794)

operations" (Holland, 2004, 315). Yet rather than thwart these public congressional activities, CIA leadership instead chose to let them become "drawn into the CIA's cover stories for PBSUCCESS" (Holland, 2004, 315). In what follows, we detail the broader role that a cover story played to motivate US overt actions.

4.3 The Cover Story Explanation

We now present support for the four novel implications outlined in Section 3.3.

4.3.1 Implication 1: Concern for strategic inferences

To parse a strategic cover story from a situation in which policymakers simply deploy a cover story after the fact, we must establish that policymakers worried ex-ante about audience attribution of covert action absent direct evidence. We find that while planning PBSUCCESS, administration officials expressed acute concerns for strategic inferences. A CIA memo in late 1953 acknowledged that "any major effort to dislodge the Communist-controlled government of Guatemala will probably be credited to the United States, and possibly on CIA" (FRUS, 1953*a*). Wisner laid out the concern more explicitly, stating that "documentary evidence may not be necessary to establish the intervention case against the United States... a strong circumstantial case could be as effective as actual evidentiary material" (CIA, 1954*e*). He went on to warn: "There is not the slightest doubt that if the operation is carried through many Latin Americans will see in it the hand of the US. But it is equally true that they would see the hand of the US in any uprising whether or not sponsored by the US" (CIA, 1954*c*). Wisner is articulating the problem that our strategic model illuminates: he understood from the outset that the *absence of evidence* of US involvement would not provide sufficiently compelling *evidence of absence* of US involvement to avoid blame for Arbenz's ouster.

4.3.2 Implication 2: Overt actions to offset strategic inferences

The position just articulated is puzzling. If policymakers worried that they would be blamed for a successful outcome even in the absence of direct evidence, and mission success required that CIA avoid attribution, then why was PBSUCCESS approved? This tension drove weeks of debate between Wisner and Henry Holland, the Assistant Secretary of State for Western Hemispheric Affairs, over whether to substantially delay the launch of PBSUCCESS in April 1954. The chance of successfully ousting Arbenz had peaked. Arbenz had been politically weakened by psychological operations, but a recent NIE suggested he was tightening controls. Thus, if CIA did not act now, they might never have the opportunity (Cullather, 2006, p42). However, Holland argued for delay because recent factors had raised the risk that US involvement would be inferred after the fact. Security officers had found recording devices in the homes of CIA staff, and Arbenz had stumbled across suggestive evidence of a US plot to fabricate a Soviet weapons cache (Cullather, 2006, p42).

In an influential memorandum summarizing this debate, Wisner began by stating the problem that needed solving. He noted that even though it "is fair to assume that no irrefutable evidence tying the project to the US Government is in the hands of the enemy", one might still worry that "if the operation is carried through many Latin Americans will see in it the hand of the US" (CIA, 1954*c*). Wisner then summarized Holland's proposed solution: the ex-post inference problem could be resolved via a moderate delay to the coup, to allow the US additional time to take overt actions. While Wisner's summary does not explain how more overt actions would resolve the ex-post inference problem, he does note, consistent with a cover story, that such a plan "would have to assume that there could be a vigorous and coordinated program of official and overt action and covert operations." Wisner instead argued in favor of moving ahead quickly without taking additional overt actions to maximize plausible deniability. In justifying his position, he noted the efficacy of actions already taken to offset strategic inferences, stating: "The security of the project is as good as can be expected and fully in keeping with the estimates made and reported on numerous occasions starting with the beginning of the project."

In subsequent weeks, Wisner's staff more directly justified overt actions as cover stories. They argued that "preliminary steps could and should start now on an official basis **maintaining fiction of no PBPRIME** [cypher for US Government] **connection PBSUCCESS**. Ambassadorial approaches could be made to Presidents in KMFLUSH [Nicaragua] and WSHOOFS [Honduras] to discuss UN/OAS issues. . . . Joint positions *must* be achieved" (CIA, 1954*g*, bolded emphasis ours, italicized in original). These covert operators explicitly acknowledged that "International politics are not for us," but they called for diplomatic action because they believed it was vital for maintaining plausible deniability given ex-post inferences.

4.3.3 Implication 3: Referring to overt actions to convince an audience

Our third implication is that administration officials should attempt to connect observed outcomes and public actions, to reduce suspicion of covert action. After Arbenz fell, and the public, media, and international community began to speculate about US involvement, we find that high-level US officials broadly publicized overt actions taken to cover up the covert actions. An NSC report, later released to the press, argued that the US contributed to pressuring Arbenz to resign through overt actions: "The Organization of American States was used as a means of achieving our objectives in the case of communist intervention in Guatemala"; when the Arbenz government appealed to the OAS to accuse its neighbors of aggression, the US "took the position that the Organization of American States was ready, willing and competent to respond to the appeal" (CIA, 1955). Similarly, in response to Congressional questioning about US covert involvement in the coup, Peurifoy pointed to the US "conducting its patrol" and "bringing the issue of the Communist threat to the hemisphere ... through legal and established channels," including the OAS (Congress, 1954, 133).

Consistent with a cover story (per Definition 1), these examples illustrate rhetorical connections made between public actions and the successful outcome *after* the outcome transpired. An implication of our strategic theory is that administration officials should also recognize the value of a cover story *before* and *during* the operation. This is harder to evidence because there is little reason to explicitly verbalize an explanation for a coup that has not yet occurred.

In our case, ex-ante opportunities for US officials to articulate a cover story arose in early 1954, when interested parties visited the US embassy seeking information about US policy toward Guatemala. In one instance, John C. Hill, Second Secretary of the US Embassy in Guatemala, recounted his conversation with a Guatemalan political elite (whose name is still classified) as follows:

> I told [redacted] that Ambassador Patterson had been quite correct in pointing out the US policy of non-intervention... but [redacted] was quite wrong in thinking that the US was not seriously concerned about the communist problem here... Assistant Secretary Cabot and others had made our concern with Communism in Guatemala abundantly clear in recent speeches; and we were now seeking means to combat Communism on

> a hemispheric basis through cooperation with other Latin American nations at the forthcoming Caracas Conference. ... *In talking in this vein to [redacted] it was my intention ... to give him the impression that the US had no concrete plan for intervention in the domestic affairs of Guatemala and continued its non-intervention policy.*[14]

This last sentence directly describes the logic of our strategic argument: the reason that Hill highlights overt policies is to disclaim involvement in covert policies *before* the coup plot had unfolded.[15]

4.3.4 Implication 4: Audience reception of the cover story

Finally, we expect that observers are naturally suspicious of US covert activity, and investigate the possibility of US involvement. But their suspicions will be offset, in part by the lack of direct evidence, and in part by their belief that US public actions contributed to ousting Arbenz.

Journalists and academics at the time analyzed US involvement in the Guatemalan affair. Two years after Arbenz was ousted, Taylor (1956)—a US-based historian of Latin America—published a comprehensive "Critique of United States Foreign Policy" surrounding Arbenz's removal. He reviewed journalistic inquiries into US policies, and academic and policy investigations into the US role published across Latin America and the United States. He also relied on interviews with confidential Guatemalan sources.

Considering the question of whether the US was directly involved in plotting the coup or training its perpetrators, Taylor finds:

> It seems clear ... that the United States did little to disabuse Arbenz' opponents of the notion that North American aid, moral and/or military, would not be lacking when the need arose. But it is difficult to find evidence which would clearly implicate Peurifoy or other United States' representatives in the plotting which resulted in Castillo's invasion. (Taylor, 1956, 793)

He separately considers whether US arms reached the revolutionaries indirectly, by way of third countries friendly to the US, and finds: "It cannot be shown that any of the arms airlifted to

[14] CIA (1954*a*); emphasis added.
[15] See CIA (1954*d*) for another similar episode.

Honduras or Nicaragua [from the United States] ultimately appeared in the hands of the Castillo Armas forces". Nonetheless, "The conclusion that the United States played an important part in the struggle in Guatemala seems inescapable"—most notably its "role in the United Nations which tended to deny to Guatemala the privileges apparently guaranteed it by its membership in that organization" (p. 797). Altogether, he concludes: "The inaction of the UN Security Council and of the Inter-American Peace Committee (as agent for the OAS) had combined with the successful operations of Castillo Armas to overthrow the Arbenz government" (p. 801).

Consistent with our theory, Taylor's inference that the US did not directly (and covertly) contribute to the revolt relies on two premises: first, that no direct evidence exists; and second, that the US was taking meaningful (and publicly observable) actions that tilted the balance in Castillo Armas's favor. Taylor's interpretation is that US overt actions and public statements contributed to Arbenz's downfall by emboldening Arbenz's opponents to mobilize against him;[16] preventing him (via the blockade) from acquiring weapons to fight back; and denying him the option to appeal to regional security institutions for support.

A broad range of investigators arrived at similar conclusions. For example, Harsch (1954) offers a more positive view on potential US intervention in Guatemala: "If there were no native revolutionary movement to encourage and support, then some other . . . remedy would have to be found." But he believes that covert action ultimately proved unnecessary because, "Fortunately, there was a bona fide native movement; and, fortunately, Honduras was willing to let it be launched from Honduran soil." Instead, the CIA meaningfully contributed by "detecting the [Soviet] shipment of arms and ammunition" and alerting the OAS to it. In a more critical account, Reston (1954) openly speculates that the CIA was involved in Guatemala, but stops short of asserting that they played a direct role in the coup—merely noting instead that the CIA was integral in uncovering the weapons cache, and exploiting that episode to foster anti-Arbenz sentiment.

We even observe similar reactions from the Arbenz government itself. In a desperate cable to the UNSC, while Castillo Armas' invasion was underway, the Guatemalan Foreign Minister accused the US of "incitement" in its public statements; of "encircling and boycotting our country", leaving Guatemala "without an air force sufficient to repel repeated acts of aggression"; and of signing

[16]See Little (2017) and Malis and Smith (2019, 2021, 2024) for theoretical work demonstrating that international messages can facilitate local coordination toward regime change.

military agreements which emboldened the "aggressor governments" of Nicaragua and Honduras (UN, 1954). Yet the cable does not accuse the US of providing direct, covert support to Castillo Armas—even though it explicitly accuses Honduras and Nicaragua of doing so.

The broader reactions at the time are consistent with our theory in two other ways. First, we assume that overt actions are costly, but less costly than the reputational harm that would follow from the revelation of covert action. The US was criticized for publicly celebrating the overthrow of a democratic government: the British Labour Party leader, for instance, was "rather shocked at the joy and approval of the American Secretary of State on the success of this *putsch*."[17] Several Latin American governments viewed the blockade as an unjustified violation of sovereignty (Friedman, 2010); but the backlash was relatively minor compared to the backlash that would have followed from PBSUCCESS' exposure.

Second, we theorize that cover stories do not fully eliminate suspicion of covert action. Rather, they offset suspicion only enough for the Intervener to avoid backlash. Indeed, some observers speculated about US involvement shortly after Arbenz fell because they understood US incentives were consistent with the outcome (e.g. Grant, 1955). But the suspicion was not enough to cause the reputational harm that had worried administration officials at the outset. Indeed, the lack of international or domestic backlash is part of the reason that historians consider PBSUCCESS a successful case of covert regime change (Immerman, 1982; Schmitz, 1999).

4.4 Alternative Explanations

We now consider the two alternative explanations detailed in Section 3.4. Policy buildups involve deploying overt actions before covert actions. As described above, Immerman (1982) comprehensively reviewed US policies before PBSUCCESS and found that the US considered but did not take overt actions in the years prior to authorizing the covert intervention (see also Cullather, 2006 35; Rabe, 1988, 43). This sequencing is sufficient to rule out the policy buildup alternative.

Comprehensive pressure involves overt actions that are expected to yield direct policy benefits that outweigh their associated costs. In Sections 4.1 and 4.2, we presented evidence that some public actions were counterproductive, or were not pursued in a manner that maximized their direct effectiveness.

[17] Quoted in Taylor (1956, 804)

Within the context of our case, two of the most plausible variants of comprehensive pressure would have induced different reactions from audiences after the mission was completed: first, that Eisenhower sought diplomatic isolation to contain Arbenz should PBSUCCESS fail to oust him; or second, that Eisenhower engaged other countries through the OAS to offset backlash in the event that PBSUCCESS was exposed. By the logic of either of these alternative explanations, audiences should have directly inferred that covert action was being taken, given the observed overt actions designed to support the covert activity. To the contrary, and consistent with the predictions of our strategic cover story, we have shown that the suspicions of prominent audiences were mitigated, rather than exacerbated, by observed overt actions.

More broadly, we have evidenced that key US decision-makers: (i) expressed an acute concern over strategic inferences; (ii) viewed plausible deniability in the face of strategic inferences to be critical to mission success—indeed, considering the mission a failure if plausible deniability was not maintained; and (iii) nonetheless authorized the covert intervention. It logically follows that these decision-makers must have found some means of managing strategic inferences, to an extent that made them willing to pursue the covert intervention. Our proposed cover story offers a coherent explanation that resolves this puzzle; alternative theories we are aware of cannot.

5 Broadening the Argument

We report two additional vignettes to illustrate our theory's applicability across different empirical contexts. Table 1 summarizes their differences. While not providing a comprehensive explanation for either case, we show that our mechanism could plausibly illuminate underappreciated dynamics that arise across diverse policy scenarios of interest to a broader range of scholars.

5.1 The Timor Gap Scandal

When East Timor gained independence from Indonesia, it inherited a maritime dispute with Australia over the oil-rich Timor Gap. Timorese leaders sought to renegotiate the existing oil concessions, which heavily favored Australia (Commonwealth of Australia, 2000). This created a vexing policy challenge for the Australian government. On the one hand, Australia viewed Timor Gap profits as an important national interest because they generated enormous tax revenue and

	Timor Gap Scandal	Cuban Missile Crisis
Leader	Australian PM Howard, FM Downer	US President Kennedy
Audience	US Gov., East Timor, Australian Public	US public, Turkey, NATO
Issue Area	Commercial negotiations, oil concessions, disputed territorial waters	Missile deployments
Open Policy	Withdraw from UNCLOS jurisdiction, delay tactics, high-priced attorneys	Public statements (audience costs), public agreement (non-invasion pledge)
Secret Policy	Bug negotiators' office to learn their reservation value	Secret diplomacy, Jupiter missile exchange
Why undesirable	Exploiting impoverished neighbor, violating international law, advancing narrow commercial interests at expense of national reputation	Appearing weak to electorate, creating moral hazard, NATO repercussions

Table 1: Summary of diverse features of cases

high-paying jobs.[18] Further, Australia worried that other neighbors might feel "aggrieved" about their own borders if the East Timor border became open to reconsideration (Pugh, 2000).

On the other hand, important international allies supported more equitable terms. Concerned with political and economic stability of the newly independent and highly oil-dependent state, over 50 US members of Congress wrote a letter calling on Australia to adhere to strict legal principles during the negotiations (Frank, 2004). Former US ambassador Peter Galbraith was appointed to negotiate on behalf of East Timor. Another issue for Australia was that East Timor could terminate the existing mining concessions, and even escalate to a formal dispute. Legal analysts believed that East Timor could accrue substantial concessions if the matter reached an international court (King, 2017).

In 2006, the parties signed the Treaty on Certain Maritime Arrangements in the Timor Sea (CMATS). While Australia made some concessions, analysts agree that CMATS substantially favored Australia (Cleary, 2007). It included a 50-50 split on the Greater Sunrise oil fields, and a commitment from East Timor that they could not renegotiate for 50 years. East Timor had privately calculated that anything less would leave them with insufficient funds to govern, and that they would be better off walking away (King, 2017, p73).

It may have seemed suspicious that Australia would extract East Timor's exact reservation

[18] It was so important that the foreign minister invoked a "national interest" exemption—used just six times in history—to fast-track the treaty's ratification (Dodd, 2007).

value. Australia attributed their success to an intensive bargaining effort. The government employed expensive outside legal consultants. They withdrew from UNCLOS's boundary dispute jurisdiction, an international treaty with broader implications, two months before East Timor's independence so that East Timor could not refer the matter to international courts (Strating, 2017). Australia also stalled profit sharing between 2003 and 2004, demanding that East Timor make important concessions. This provoked backlash from American elites, who argued that Australia was taking advantage of their neighbor's impoverished position (Frank, 2004). Still, in 2006, East Timor accepted CMATS, believing that Australia had adhered to international law during negotiations.

This was not the case. Secretly, the Australian Secret Intelligence Service had illegally bugged the offices of East Timor's prime minister and key negotiators (Cannane, 2015). Thus, unbeknownst to Galbraith and the Timorese, the Australians knew exactly what their counterparts were willing to accept. Australia's secret efforts were exposed by a whistleblower, Witness K, who came forward once he learned that former Australian Foreign Minister Alexander Downer was given a lucrative advisory position with Woodside Petroleum, the firm that profited the most from the episode.

The reaction of U.S. elites is consistent with our characterization of audiences concerned with preventing unscrupulous policy actions. Galbraith, as a former ambassador with "frequent access to US intelligence," had "never…seen his country attempt an operation as commercially driven as Australia's was." Galbraith described the measure as "outrageous…It was not what you do to a friendly state. And it was not something you do for commercial advantage…the Howard and Downer government, they were shills for the corporations" (Knaus, 2019).

Notably, Australia took steps during negotiations to conceal their illicitly obtained private knowledge. They did not demand large concessions on the first day. Rather, over a series of weeks they carefully crafted arguments to arrive at the final position (Knaus, 2019). We suggest that this strategy, along with the aforementioned public actions, contributed to a cover story.

5.2 The Cuban Missile Crisis

According to public understanding at the time, the Cuban Missile Crisis ended when the USSR withdrew missiles from Cuba in return for a vague commitment from the US not to invade Cuba. It is now well known that the resolution of the crisis is largely attributable to Kennedy's secret commitment to remove Jupiter missiles from Turkey (see Criss, 1997). The Kennedy Administration insisted on secrecy because they were concerned about the political fallout at home and abroad should the quid pro quo become public (Bernstein, 1980). Robert Kennedy, for instance, refused to deliver a letter about the exchange from Khrushchev to the president, "since who knows where and when such letters can surface... The appearance of such a document could cause irreparable harm to my political career in the future" (National Security Archive, 1962).

Many argue that the official deal was so lopsided that it raised suspicions that something else was going on (Scott and Hughes, 2015, p173). Even at the time, many speculated that a missile exchange could have facilitated peace;[19] indeed, Khrushchev and others had raised the possibility during the crisis.

How did Kennedy offset this suspicion? Scholars have emphasized the extraordinary secret efforts that the Kennedy Administration took to disclaim a connection between removing Jupiter missiles and the Cuban Crisis (Scott and Hughes, 2015; Bernstein, 1980). Kennedy denied the missile exchange, even in private conversations with former Presidents Eisenhower and Truman. Kennedy's inner circle also vilified UN Ambassador Adlai Stevenson, who was the earliest and most prominent advocate for the missile exchange during the crisis. Kennedy further minimized suspicion by waiting five months to remove the missiles from Turkey, and by removing missiles from Italy and Turkey simultaneously, to give the appearance of a broader effort to restructure forces.

Our theory sheds light on two underappreciated aspects of this case. First, it suggests that the non-invasion pledge played a more important role in advancing this fiction than scholars appreciate because it gave Kennedy a fall-back position. Indeed, when Truman asked Kennedy directly if Kennedy had made a missile exchange, Kennedy replied, "they came back with and accepted the earlier proposal" on the non-invasion pledge (quoted in Stern, 2003).

Second, we illuminate the *indirect* effect of audience costs (Fearon, 1994). Under the standard logic, leaders (Kennedy) use aggressive public statements to convince rivals (here the Soviets) that they will not back down, which should engender concessions from the rival. But in this case, alongside his public statements of resolve, Kennedy secretly offered the Soviets a substantial concession. This raises the question: why did Kennedy make these public statements at all?

We suggest that the ostensible generation of audience costs could function as a cover story. By this interpretation, Kennedy did not necessarily intend to convince the Soviets that he would

[19] Sufficient evidence for academic speculation emerged during the 1970s (Allison, 1971; Bernstein, 1976).

escalate if they did not capitulate. Rather, to raise the credibility of the official line, he needed to convince outside observers that the Soviets did capitulate for fear of escalation. Indeed, the Administration used their tough public stance to help explain why the Soviets eventually backed down, with Secretary of State Rusk famously stating: "We're eyeball to eyeball, and I think the other fellow just blinked."

6 Discussion

When a government holds both secret and overt policy levers, and direct evidence of secret policies is hard to uncover, audiences develop rational suspicion that secret policies are being pursued. How can a government achieve its policy objectives and offset this suspicion? We argue that cover stories provide an underappreciated tool for reputational management in the face of strategic inferences. We provide an in-depth examination of the use of a cover story in a U.S. covert regime change operation, and demonstrate that the same mechanism is plausibly at work in other cases of commercial negotiations and nuclear brinkmanship.

The general strategic problem that we study manifests across a wide variety of substantive domains in American, comparative, and international politics. We conclude by briefly discussing how our framework may be fruitfully applied to longstanding debates across these research areas.

Autocratic elections. Scholars have shown that autocratic governments frequently enact seemingly popular and responsive public policies in advance of elections which they will ultimately rig (Magaloni, 2006; Simpser, 2013; Miller, 2015*a*,*b*; Rød, 2019; Han, 2022). Our theory suggests that this dual approach is valuable in part due to domestic and international audiences' uncertainty over the true cause of a strong electoral performance: the overt policies provide a cover story which can reduce ex-post suspicion of electoral fraud.

Money in politics. When special interests expend vast resources on lobbying government officials, are they engaging in persuasion and information transmission, or more nefarious quid-pro-quo exchange (Grossman and Helpman, 2001; Tripathi, Ansolabehere and Snyder Jr, 2002; Kalla and Broockman, 2016; McKay, 2020; Schnakenberg and Turner, 2024; Furnas, LaPira and Brock, 2025; Kim, Stuckatz and Wolters Freiheyt, 2026)? One interpretation is that the overt activities lobbyists engage in—conducting policy research, preparing and disseminating reports, public advocacy, etc.—can serve as a cover story to explain why a policymaker came to adopt the lobbyist's preferred position.

Biases and favoritism in hiring. Across the private sector (Uhlmann and Cohen, 2005; Castilla and Benard, 2010; Rivera, 2012), public bureaucracies (Freedman, 1988; Bagues and Esteve-Volart, 2010; Greenan, Lanfranchi, l'Horty, Narcy and Pierné, 2019; Perez-Chiques and Rubin, 2022), and, notably, in academia (Combes, Linnemer and Visser, 2008; Durante, Labartino and Perotti, 2011; Zinovyeva and Bagues, 2015; Fisman, Shi, Wang and Xu, 2018), hiring and promotion decisions are ostensibly governed by neutral, rules-based procedures; yet these same processes often select candidates whom the selectors were predisposed to favor for personal, political, or demographic reasons. The publicly espoused selection process can serve as a cover story for the actual means by which the chosen candidate emerged: outsiders cannot readily determine whether the favored candidate would have prevailed under a genuinely neutral process.

Technology acquisition. Some firms and governments simultaneously engage in industrial espionage and intellectual property theft, on the one hand, while pursuing publicized R&D campaigns and touting their own ingenuity on the other (Ben-Atar, 2004; Hannas, Mulvenon and Puglisi, 2013; Lindsay, 2014; Gilli and Gilli, 2018; Glitz and Meyersson, 2020). The latter set of activities may provide a cover story to explain innovation success while disclaiming any involvement in the former.

Finally, our theory holds policy implications as US-China competition intensifies while US legitimacy still relies on liberal values (Joseph, 2026). Past scholars have suggested that the US could exploit tight mission controls to avoid the normatively damaging direct exposure of covert action but succeed in competition with normatively less-constrained rivals (Poznansky, 2025). Our framework shows that strategic inferences place a practical limit on how often the US can avoid suspicion in the absence of evidence of any policy, but overt cover stories may ease suspicion that the US succeeds via hypocritical covert actions for longer.

References

ACHRE. 1996. "Final Report (061-000-00-848-9).".

ADST. 2023. "Oral Histories Project, Cuba." Association for Diplomatic Studies and Training. https://adst.org/Readers/Cuba.pdf.

Allison, Graham. 1971. *Essence of Decision: Explaining the Cuban Missile Crisis.* Little, Brown and Company.

Ashworth, Scott. 2012. "Electoral Accountability: Recent Theoretical and Empirical Work." *Annual Review of Political Science* 15:183–201.

Axelrod, R. and R. Iliev. 2014. "Timing of cyber conflict." *Proceedings of the National Academy of Sciences* 111:1298–1303.

Backus, Matthew and Andrew T. Little. 2020. "I Don't Know." *American Political Science Review* 114(3):724–743.

Bagues, Manuel F and Berta Esteve-Volart. 2010. "Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment." *The Review of Economic Studies* 77(4):1301–1328.

Baliga, Sandeep, Ethan Bueno de Mesquita and Alexander Wolitzky. 2020. "Deterrence with Imperfect Attribution." *American Political Science Review* 114:1155–1178.

Bates, Robert H, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry R Weingast. 1998. *Analytic narratives.* Princeton University Press.

Ben-Atar, Doron S. 2004. *Trade secrets: Intellectual piracy and the origins of American industrial power.* Yale University Press.

Bernstein, Barton J. 1976. "The week we almost went to war." *Bulletin of the Atomic Scientists* 32:12–21. doi: 10.1080/00963402.1976.11455563.

Bernstein, Barton J. 1980. "The Cuban Missile Crisis: Trading the Jupiters in Turkey?" *Political Science Quarterly* 95:97–125.

Bils, Peter and Bradley C Smith. 2025. "The logic of secret alliances." *American Journal of Political Science* .

Bloch, Chase and Roseanne W. McManus. 2024. "Denying the Obvious: Why Do Nominally Covert Actions Avoid Escalation?" *International Organization* 78:600–624.

Bowen, Gordon L. 1983. "U.S. Foreign Policy toward Radical Change: Covert Operations in Guatemala, 1950-1954." *Latin American Perspectives* 10(1):88–102.

Brewer, Sam Pope. 1961*a*. "EXILES SAY CUBA SEEKS 'INCIDENT'; Castro Said to Be Trying to Provoke Crisis, Even Intervention by U.S.". The New York Times; January 7, 1961. https://www.nytimes.com/1961/01/07/archives/exiles-say-cuba-seeks-incident-castro-said-to-be-trying-to-provoke.html.

Brewer, Sam Pope. 1961*b*. "Exiles Say Military Resistance to Castro Is Growing Inside Cuba.". The New York Times; January 14, 1961. https://www.nytimes.com/1961/01/14/archives/exiles-say-military-resistance-to-castro-is-growing-inside-cuba.html.

Canes-Wrone, Brandice, Michael C. Herron and Kenneth W. Shotts. 2001. "Leadership and Pandering: A Theory of Executive Policymaking." *American Journal of Political Science* 45:532–550.

Canfil, Justin Key. 2022. "The illogic of plausible deniability: why proxy conflict in cyberspace may no longer pay." *Journal of Cybersecurity* 8.

Cannane, Steve. 2015. "'Matter of death and life': Espionage in East Timor and Australia's diplomatic bungle." *ABC News* .

Carnegie, Allison. 2021. "Secrecy in International Relations and Foreign Policy." *Annual Review of Political Science* 24:213–233.

Carson, Austin. 2018. *Secret wars : covert conflict in international politics*. Princeton University Press.

Castilla, Emilio J and Stephen Benard. 2010. "The paradox of meritocracy in organizations." *Administrative science quarterly* 55(4):543–676.

CIA. 1953. "Wisner's Memorandum for Chief, Western Hemisphere Division (CREST, No. 0000914589).". https://www.cia.gov/readingroom/docs/DOC_0000914589.pdf.

CIA. 1954*a*. "Dispatch No. HGG-A-619. Memorandum of Conversation, From COS, Guatemala to Chief WHD (CREST DOC˙0000914259).". https://www.cia.gov/readingroom/docs/DOC_0000914259.pdf.

CIA. 1954*b*. "Memorandum for Col. J. C. King. Doc No. 0000914115.". https://www.cia.gov/readingroom/docs/DOC_0000914115.pdf.

CIA. 1954*c*. "Memorandum From the Deputy Director for Plans of the Central Intelligence Agency (Wisner) to Director of Central Intelligence Dulles." *Central Intelligence Agency, Job 79–01025A, Box 151, Folder 6. Top Secret.* . https://history.state.gov/historicaldocuments/frus1952-54Guat/d133.

CIA. 1954*d*. "Memorandum of Conversation from John Hill to The Chargé d'Affaires a.i., Jan. 25, 1954. (Dispatch HGG-A-689).". https://www.cia.gov/readingroom/docs/DOC_0000913985.pdf.

CIA. 1954*e*. "Minutes of Weekly PBSUCCESS Meeting. 20 April, 1954 (CREST No. 0000916252).". https://www.cia.gov/readingroom/docs/DOC_0000916252.pdf.

CIA. 1954*f*. "Synthesis of [classified]'s remarks relevant to PBSUCCESS. April 22, 1954 (CREST no. DOC˙0000135899).". https://www.cia.gov/readingroom/docs/DOC_0000135899.pdf.

CIA. 1954*g*. "Telegram From Operation PBSUCCESS Headquarters in Florida to the Central Intelligence Agency." *Central Intelligence Agency, Job 79–01025A, Box 6, Folder 3.* . https://history.state.gov/historicaldocuments/frus1952-54Guat/d218.

CIA. 1955. "Progress Report on NSC 5432/1 (Crest No. CIA-RDP80R01731R003000030008-2).".

CIA. 1960. "NIE 85-2-60, the Situation in Cuba (CREST no. 0000132646).". https://www.cia.gov/readingroom/docs/DOC_0000132646.pdf.

Cleary, Paul. 2007. *Shakedown: Australia's Grab for Timor Oil.* Allen & Unwin.

Coe, Andrew J. 2018. "Containing Rogues: A Theory of Asymmetric Arming." *The Journal of Politics* 80:1197–1210.

Colaresi, Michael. 2012. "A Boom with Review: How Retrospective Oversight Increases the Foreign Policy Ability of Democracies." *American Journal of Political Science* 56:671–689.

Combes, Pierre-Philippe, Laurent Linnemer and Michael Visser. 2008. "Publish or peer-rich? The role of skills and networks in hiring economics professors." *Labour Economics* 15(3):423–441.

Commonwealth of Australia. 2000. "Final report on the inquiry into East Timor (Ch 4).".

Congress. 1954. "Communist aggression in Latin America. : Ninth interim report of hearings before the Subcommittee on Latin America of the Select Committee on Communist Aggression, House of Representatives, Eighty-third Congress, second session, under authority of H. Res. 346 and H. Res. 438. Guatemala.". https://catalog.hathitrust.org/Record/100662563.

Criss, Nur Bilge. 1997. "Strategic nuclear missiles in Turkey: The Jupiter affair, 1959–1963." *Journal of Strategic Studies* 20:97–122. doi: 10.1080/01402399708437689.

Cullather, N. 2006. *Secret History: The CIA's Classified Account of Its Operations in Guatemala, 1952-1954.* Stanford University Press.

Debs, Alexandre and Nuno P Monteiro. 2014. "Known unknowns: Power shifts, uncertainty, and war." *International Organization* 68(1):1–31.

Dodd, Mark. 2007. "Downer rushes on Timor deal." *The Australian* .

DoS. 2023. "A Guide to the United States' History of Recognition, Diplomatic, and Consular Relations, by Country, since 1776: Cuba.". https://history.state.gov/countries/cuba.

Durante, Ruben, Giovanna Labartino and Roberto Perotti. 2011. Academic dynasties: decentralization and familism in the Italian academia. Technical report National Bureau of Economic Research.

Eisenhower, Dwight. 1960. "Statement by the President Upon Issuing Proclamation Fixing the Cuban Sugar Quota at Zero.".
URL: *https://www.presidency.ucsb.edu/documents/statement-the-president-upon-issuing-proclamation-fixing-the-cuban-sugar-quota-zero*

Fearon, J. D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88:577–592.

Fisman, Raymond, Jing Shi, Yongxiang Wang and Rong Xu. 2018. "Social ties and favoritism in Chinese science." *Journal of Political Economy* 126(3):1134–1171.

Frank, Barney. 2004. "Letter to John Howard.". https://www.etan.org/news/2004/03houseltr.htm#letter.

Fraser, Andrew. 2005. "Architecture of a broken dream: The CIA and Guatemala, 1952–54." *Intelligence and National Security* 20(3):486–508.

Freedman, Anne. 1988. "Doing battle with the patronage army: Politics, courts, and personnel administration in Chicago." *Public Administration Review* pp. 847–859.

Friedman, Max Paul. 2010. "Fracas in Caracas: Latin American Diplomatic Resistance to United States Intervention in Guatemala in 1954." *Diplomacy and Statecraft* 21(4):669–689.

FRUS. 1953*a*. "Draft Memorandum for the Record: Program For PBSUCCESS, November 12, 1953.". https://history.state.gov/historicaldocuments/frus1952-54Guat/d65.

FRUS. 1953*b*. "Letter From the Ambassador to Guatemala (Peurifoy) to the Assistant Secretary of State for Inter-American Affairs (Cabot).". https://history.state.gov/historicaldocuments/frus1952-54Guat/d79.

FRUS. 1954. "Memorandum for the Record. March 14.". https://history.state.gov/historicaldocuments/frus1952-54Guat/d116.

FRUS. 1960. "Memorandum of Discussion at the 450th Meeting of the National Security Council, Washington, July 7, 1960.". https://history.state.gov/historicaldocuments/frus1958-60v06/d545.

Furnas, Alexander C, Timothy M LaPira and Clare Brock. 2025. "Conflating lobbying and PACs: The surprisingly low overlap in organizational lobbying and campaign expenditures." *PS: Political Science & Politics* 58(4):597–605.

Gerring, John. 2004. "What Is a Case Study and What Is It Good for?" *American Political Science Review* 98:341–354.

Gilli, Andrea and Mauro Gilli. 2018. "Why China has not caught up yet: Military-technological superiority and the limits of imitation, reverse engineering, and cyber espionage." *International Security* 43(3):141–189.

Glitz, Albrecht and Erik Meyersson. 2020. "Industrial espionage and productivity." *American Economic Review* 110(4):1055–1103.

Goemans, Hein and William Spaniel. 2016. "Multimethod Research: A Case for Formal Theory." *Security Studies* 25(1):25–33.

Grant, Donald. 1955. "Guatemala and the United States' Foreign Policy." *Journal of International Affairs* 9:64–72.

Greenan, Nathalie, Joseph Lanfranchi, Yannick l'Horty, Mathieu Narcy and Guillaume Pierné. 2019. "Do competitive examinations promote diversity in civil service?" *Public Administration Review* 79(3):370–382.

Grossman, Gene M. and Elhanan Helpman. 2001. *Special Interest Politics*. Cambridge, MA: MIT Press.

Han, Kangwook. 2022. "Political budgetary cycles in autocratic redistribution." *Comparative Political Studies* 55(5):727–756.

Hannas, William C, James Mulvenon and Anna B Puglisi. 2013. *Chinese industrial espionage: Technology acquisition and military modernisation*. Routledge.

Harsch, Joseph. 1954. "State of the Nations Guatemalan Interlude." *Christian Science Monitor* .

Holland, Max. 2004. "Operation PBHISTORY: The Aftermath of SUCCESS." *International Journal of Intelligence and CounterIntelligence* 17(2):300–332.
URL: *https://doi.org/10.1080/08850600490274935*

Horowitz, Michael. 2021. "Statement of Michael E. Horowitz, Inspector General, U.S. Department of Justice before the U.S. House of Representatives Committee on Oversight and Reform concerning "Accountability and Lessons Learned from the Trump Administration's Child Separation Policy".".

Immerman, Richard. 1982. *The CIA in Guatemala.* University of Texas Press.

Izzo, Federica. 2023. "Ideology for the Future." *American Political Science Review* 117(3):1089–1104.

Jeffreys-Jones, Rhodri. 2022. *A Question of Standing: The History of the CIA.* Oxford University Press.

Joseph, Michael. 2026. *The Origins of Great Power Rivalries.* Cambridge University Press.

Joseph, Michael F and Michael Poznansky. 2018. "Media technology, covert action, and the politics of exposure." *Journal of Peace Research* 55:320–335.

Joseph, Michael F and Michael Poznansky. 2024. "Secret innovation." *International Organization* 78(4):766–799.

Joseph, Michael F., Michael Poznansky and William Spaniel. 2022. "Shooting the Messenger: The Challenge of National Security Whistleblowing." *The Journal of Politics* 84(2):846–860.

Judd, Gleason. 2017. "Showing Off: Promise and Peril in Unilateral Policymaking." *Quarterly Journal of Political Science* 12(2):241–268.

Kalla, Joshua L and David E Broockman. 2016. "Campaign contributions facilitate access to congressional officials: A randomized field experiment." *American Journal of Political Science* 60(3):545–558.

Kim, In Song, Jan Stuckatz and Lukas Wolters Freiheyt. 2026. "Systemic and sequential links between campaign donations and lobbying." *The Journal of Politics* 88(1):110–124.

King, Robert J. 2017. "The Timor Gap, 1972–2017." Submission to the Parliament of Australia, Certain Maritime Arrangements—Timor-Leste, Submission 27.

Knaus, Christopher. 2019. "Witness K and the 'outrageous' spy scandal that failed to shame Australia." *The Guardian* .

Kurizaki, Shuhei. 2007. "Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy." *American Political Science Review* 101:543–558.

Levi, Margaret and Barry R. Weingast. 2022. Analytic Narratives and Case Studies. In *The Case for Case Studies.* Cambridge University Press pp. 239–257.

Levin, Dov H. 2020. *Meddling in the ballot box: The causes and effects of partisan electoral interventions.* Oxford University Press.

Lindsay, Jon R. 2014. "The impact of China on cybersecurity: Fiction and friction." *International Security* 39(3):7–47.

Little, Andrew T. 2017. "Coordination, Learning, and Coups." *Journal of Conflict Resolution* 61:204–234.

Lorentzen, Peter, M Taylor Fravel and Jack Paine. 2017. "Qualitative investigation of theoretical models: the value of process tracing." *Journal of Theoretical Politics* 29:467–491.

Magaloni, Beatriz. 2006. *Voting for Autocracy: Hegemonic Party Survival and Its Demise in Mexico*. Cambridge: Cambridge University Press.

Malis, Matt and Alastair Smith. 2019. "A global game of diplomacy." *Journal of Theoretical Politics* 31(4):480–506.

Malis, Matt and Alastair Smith. 2021. "State visits and leader survival." *American Journal of Political Science* 65(1):241–256.

Malis, Matt and Alastair Smith. 2024. "Quid Pro Quo Diplomacy." *Games* 15(2):14.

Maskin, Eric and Jean Tirole. 2004. "The politician and the judge: Accountability in government." *American Economic Review* 94(4):1034–1054.

McKay, Amy Melissa. 2020. "Buying amendments? Lobbyists' campaign contributions and microlegislation in the creation of the Affordable Care Act." *Legislative Studies Quarterly* 45(2):327–360.

Miller, Michael K. 2015*a*. "Elections, information, and policy responsiveness in autocratic regimes." *Comparative Political Studies* 48(6):691–727.

Miller, Michael K. 2015*b*. "Electoral authoritarianism and human development." *Comparative Political Studies* 48(12):1526–1562.

Myrick, Rachel. 2020. "Why So Secretive? Unpacking Public Attitudes toward Secrecy and Success in US Foreign Policy." *The Journal of Politics* 82:828–843.

National Security Archive. 1962. "Soviet Embassy, Telegram, Report from Ambassador Dobrynin on Meeting with Robert Kennedy Regarding Khrushchev letter of October 28, 1962, October 30, 1962." George Washington University. https://nsarchive.gwu.edu/document/29365-document-4-soviet-embassy-telegram-report-ambassador-dobrynin-meeting-robert-kennedy.

New York Times. 1961. "U.S. Pro-Castro Unit Asks Inquiry on C.I.A.". January 6, 1961.
URL: *https://www.nytimes.com/1961/01/06/archives/us-procastro-unit-asks-inquiry-on-cia.html*

Perez-Chiques, Elizabeth and Ellen V Rubin. 2022. "Debasement of merit: The method and experience of political discrimination by public employees in the commonwealth of Puerto Rico." *Review of Public Personnel Administration* 42(4):669–685.

Poznansky, Michael. 2019. "Feigning compliance: covert action and international law." *International Studies Quarterly* 63(1):72–84.

Poznansky, Michael. 2020. *In the shadow of international law : secrecy and regime change in the postwar world*.

Poznansky, Michael. 2022. "Revisiting plausible deniability." *Journal of Strategic Studies* 45:511–533.

Poznansky, Michael. 2025. *Great Power, Great Responsibility: How the Liberal International Order Shapes U.S. Foreign Policy.* Oxford University Press.

Pugh, Wendy. 2000. "Australia seeks to avoid East Timor border dispute." *Reuters* .

Rabe, Stephen G. 1988. *Eisenhower and Latin America: the foreign policy of anticommunism.* University of North Carolina Press.

Reston, James. 1954. "Washington: With the Dulles Brothers in Darkest Guatemala." *New York Times* .

Rivera, Lauren A. 2012. "Hiring as cultural matching: The case of elite professional service firms." *American sociological review* 77(6):999–1022.

Rød, Espen Geelmuyden. 2019. "Fraud, grievances, and post-election protests in competitive authoritarian regimes." *Electoral Studies* 58:12–20.

Schlesinger, S C and S Kinzer. 1982. *Bitter Fruit: The Untold Story of the American Coup in Guatemala.* Anchor books Doubleday.

Schmitz, David F. 1999. *Thank God They're on Our Side: The United States and Right-wing Dictatorships, 1921-65.* University of North Carolina Press.

Schnakenberg, Keith E and Ian R Turner. 2024. "Formal theories of special interest influence." *Annual Review of Political Science* 27(1):401–421.

Scott, L and R G Hughes. 2015. *The Cuban Missile Crisis: A Critical Reappraisal.* Taylor and Francis.

Simpser, Alberto. 2013. *Why governments and parties manipulate elections.* Cambridge University Press.

Smith, Gregory L. 2019. "Secret but constrained: the impact of elite opposition on covert operations." *International Organization* 73(3):685–707.

Spaniel, William and Michael Poznansky. 2018. "Credible Commitment in Covert Affairs." *American Journal of Political Science* 62:668–681.

Stern, Sheldon. 2003. *Averting 'The Final Failure': John F. Kennedy and the Secret Cuban Missile Crisis Meetings.* Stanford University Press.

Strating, Rebecca. 2017. "Law of the Sea: Settling the Australia and Timor-Leste Dispute." Australian Institute of International Affairs.

Taylor, Philip B. 1956. "The Guatemalan Affair: A Critique of United States Foreign Policy." *American Political Science Review* 50:787–806.

Tripathi, Micky, Stephen Ansolabehere and James M Snyder Jr. 2002. "Are PAC contributions and lobbying linked? New evidence from the 1995 Lobby Disclosure Act." *Business and politics* 4(2):131–155.

Uhlmann, Eric Luis and Geoffrey L Cohen. 2005. "Constructed criteria: Redefining merit to justify discrimination." *Psychological science* 16(6):474–480.

UN. 1954. "Cablegram dated 19 June 1954 from the Minister for External Relations of Guatemala addressed to the President of the Security Council. (S3232).". https://digitallibrary.un.org/record/602948.

Wallace, Geoffrey PR. 2024. *International law and the public: how ordinary people shape the global legal order.* Cornell University Press.

Yoder, Brandon K. and William Spaniel. 2022. "Costly Concealment: Secret Foreign Policymaking, Transparency, and Credible Reassurance." *International Organization* 76:868–900.

Zinovyeva, Natalia and Manuel Bagues. 2015. "The role of connections in academic promotions." *American Economic Journal: Applied Economics* 7(2):264–292.

7 Appendix

7.1 Non-Strategic Model of Audience Learning: Formal Results

Here we provide the full formal characterization of the non-strategic model of audience learning presented in Section 2. Recall that the government can take public action $a_p \in \{0, 1\}$, and/or covert action $a_c \in \{0, 1\}$. The government's policy profile $a = (a_p, a_c) \in \{0, 1\}^2$ is randomly generated according to the following commonly known probabilities:

$$\begin{aligned} \nu_p &= Pr(a_p = 1, a_c = 0) \\ \nu_c &= Pr(a_p = 0, a_c = 1) \\ \nu_{pc} &= Pr(a_p = 1, a_c = 1) \\ \nu_0 &= Pr(a_p = 0, a_c = 0) = 1 - \nu_p - \nu_c - \nu_{pc} \end{aligned} \tag{5}$$

The actions then contribute to either a policy success ($y = 1$) or policy failure ($y = 0$), as follows:

$$Pr(y = 1|a_p, a_c) = \begin{cases} \alpha_0, & a_p = 0, a_c = 0 \\ \alpha_p, & a_p = 1, a_c = 0 \\ \alpha_c, & a_p = 0, a_c = 1 \\ \alpha_{pc}, & a_p = 1, a_c = 1 \end{cases} \tag{6}$$

where $\alpha_c > \alpha_0$, $\alpha_p > \alpha_0$, and $\alpha_{pc} = \alpha_p + (1 - \alpha_p)\alpha_c$. We note that this specification of α_{pc} allows for the policy success function to either be supermodular (public and covert action are complements) or submodular (public and covert action are substitutes) in the inputs. To see this, observe that the effect of covert action when public action is *not* being used is $\alpha_c - \alpha_0$; and the effect of covert action when public action *is* being used is $\alpha_{pc} - \alpha_p = \alpha_c(1 - \alpha_p)$. Thus the actions are complements if $\alpha_0 > \alpha_p\alpha_c$, and substitutes if $\alpha_0 < \alpha_p\alpha_c$.

The audience observes both the public action and the policy outcome with certainty. Let $z \in \{0, 1\}$ denote whether the audience observes direct evidence of the government's use of covert action; this is generated according to the probability $Pr(z = 1|a_c) = a_c\lambda$, for $\lambda \in [0, 1]$.

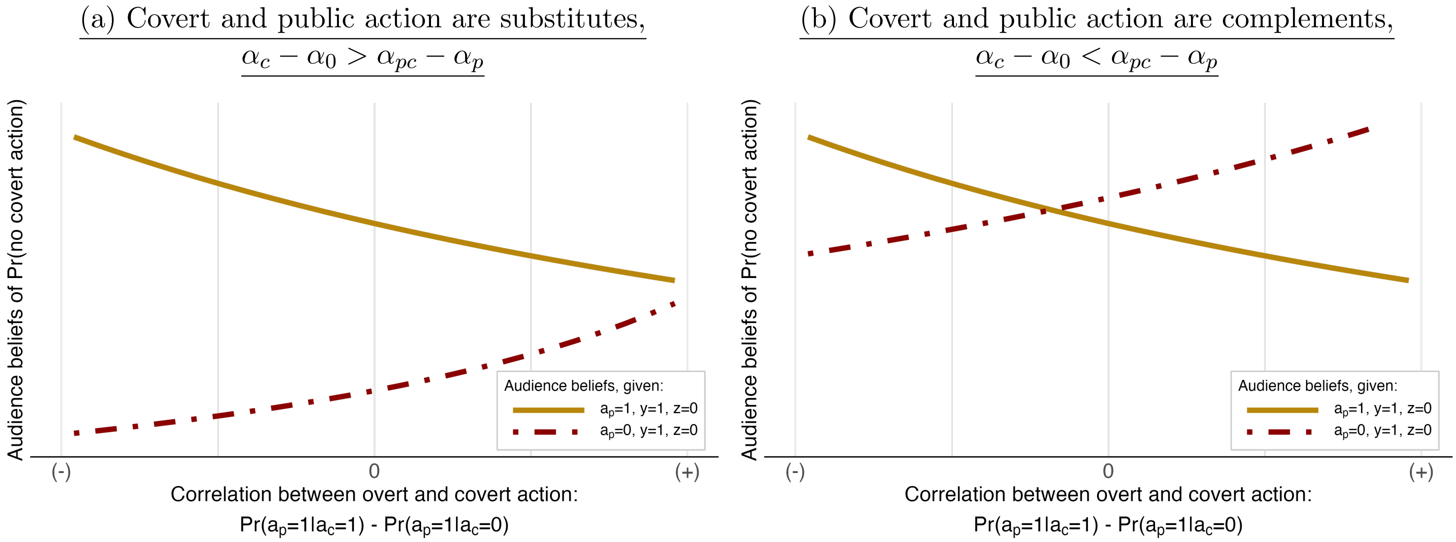
Then the audience's posterior beliefs, $\eta^{a_p,y,z} = Pr(a_c = 0|a_p, y, z)$, are as follows:

$$\eta^{a_p,y,z=0} = \frac{Pr(y|a_c = 0, a_p)Pr(a_c = 0, a_p)}{Pr(y|a_c = 0, a_p)Pr(a_c = 0, a_p) + (1 - \lambda)Pr(y|a_c = 1, a_p)Pr(a_c = 1, a_p)}$$

$$\eta^{1,1,0} = \frac{\alpha_p\nu_p}{\alpha_p\nu_p + (1 - \lambda)\alpha_{pc}\nu_{pc}} = \frac{1}{1 + (1 - \lambda)\frac{\alpha_{pc}\nu_{pc}}{\alpha_p\nu_p}}$$

$$\eta^{1,0,0} = \frac{1}{1 + (1 - \lambda)\frac{(1-\alpha_{pc})\nu_{pc}}{(1-\alpha_p)\nu_p}}, \qquad \eta^{0,1,0} = \frac{1}{1 + (1 - \lambda)\frac{\alpha_c\nu_c}{\alpha_0\nu_0}}, \qquad \eta^{0,0,0} = \frac{1}{1 + (1 - \lambda)\frac{(1-\alpha_c)\nu_c}{(1-\alpha_0)\nu_0}}$$

Figure 4: Comparing audience beliefs, given interactive effects of covert and public action



Altogether, this brings us to a formal statement and proof of Result 1:

Proof of Result 1: Public action functions as a cover story if and only if $\eta^{1,1,0} > \eta^{0,1,0}$, which rearranges to

$$\frac{\alpha_0 \alpha_{pc}}{\alpha_p \alpha_c} < \frac{\nu_p \nu_c}{\nu_0 \nu_{pc}}$$

Recall from above that covert and public action are (weak) substitutes iff $\alpha_0 \leq \alpha_p \alpha_c$. Then observe that $\frac{\nu_p \nu_c}{\nu_0 \nu_{pc}} \geq 1 \iff Pr(a_p = 1|a_c = 1) \leq Pr(a_p = 1|a_c = 0)$, which holds iff the uses of public and covert action are (weakly) negatively correlated.

Altogether we obtain the following:

- If covert and public action are (weak) substitutes, and their use is weakly negatively correlated, then $\frac{\alpha_0 \alpha_{pc}}{\alpha_p \alpha_c} \leq \alpha_{pc} \leq 1 \leq \frac{\nu_p \nu_c}{\nu_0 \nu_{pc}}$. If at least one of these inequalities is strict, then public action functions as a cover story.
- If $\alpha_{pc} < 1$, there exists a non-empty range of parameters for which covert and public action are both complements and positively correlated, and public action still functions as a cover story.

■

Figure 4 provides two alternative versions of Figure 1. The main text figure featured additive effects of covert and public action, while the two cases presented here depict the cases of complements and substitutes. All three cases are generated with the parameter values $\nu_0 = \nu_p = 0.25$; ν_c ranging from 0 to 0.5 (and $\nu_{pc} = 1 - \nu_c - \nu_p - \nu_0$); $\lambda = 0.5$; $\alpha_p = \alpha_c = 0.5$; and $\alpha_{pc} = \alpha_p + (1 - \alpha_p)\alpha_c = 0.75$. In the additive effects case, we have $\alpha_0 = \alpha_p + \alpha_c - \alpha_{pc} = 0.25$. In the first panel of Figure 4, with substitutes, we have $\alpha_0 = 0.1$. In the second panel, with complements, we have $\alpha_0 = 0.4$.

We label the horizontal axis as representing variation in the correlation between overt and covert action. More precisely, the horizontal axis captures the difference in probabilities of public action conditional on covert action vs. covert inaction, $Pr(a_p = 1|a_c = 1) - Pr(a_p = 1|a_c = 0)$. This is equivalent to the coefficient from a bivariate regression of public action on covert action, which equals the correlation between covert and public action rescaled by $\sqrt{V[a_p]/V[a_c]}$.

7.2 Comments on Modeling Assumptions

Here we provide additional discussion and justification of some assumptions of the strategic model introduced in the main text.

Privately-known costs of covert action: The model assumes that the leader's cost of covert action, k_c, is privately known to the leader (and not observed by the audience). This is a natural assumption, given that the covert action itself is taken in secret. Having k_c be private information is not necessary for our results; the alternative would involve L playing a mixed covert action strategy, which proves more technically unwieldy than our current setup. In our analysis, if the covert action is exposed, it makes no difference whether the audience also observes k_c.

Additionally, note that we model A's uncertainty over the *cost* of covert action (rather than its *effectiveness*) for technical simplicity. Revising this assumption would not substantially alter the results. However, assuming uncertainty over the effectiveness of *public* action is important for generating the model's core substantive insights.

False positives: We assume that the probability that the audience observes direct evidence of covert action is $Pr(z = 1|a_c) = a_c\lambda$. This means that the audience does not observe "false positive" revelations (i.e., what they believe to be "direct evidence" of covert action, even though covert action was not taken). This assumption helps us to better disentangle the conceptual issues surrounding inferences drawn from direct versus circumstantial evidence. Allowing for some probability of false positive signals ($Pr(z = 1|a_c = 0) > 0$) would further complicate the audience's inferential challenge. The core substantive insights of our analysis would remain unchanged as long as the probability of false positives is sufficiently low; we defer a complete analysis of this possibility to future research.

Audience's objective function: Our model assumes that the audience has direct preferences over the type of leader that they punish or reward, $U_A = \mathbb{1}[r = \theta]$. We could instead consider that after the audience's punishment/reward decision, there is a second period of policymaking, which proceeds analogously to the first (with new realizations of k_c and ω). If the audience plays $r = 1$, the initial leader sets policy in the second period; if $r = 0$, a new leader is selected. (Following the discussion in the main text, this action can represent the domestic electorate's choice over whether to retain the leader, or a foreign government's choice over whether to continue cooperating with the leader.) Suppose the audience receives some positive utility from a successful policy outcome, and some disutility from the leader's use of covert action: $U_A = (\gamma)y^{t=2} + (1 - \gamma)(-a_c^{t=2}k_c^{t=2})$ for some $\gamma \in (0, 1)$. Then, as long as γ is not too large, all results from our current setup remain substantively unchanged.

Audience's punishment strategy: The main text analysis shows that the audience uses a "punishment threshold" of $\frac{1}{2}$. This arises from the assumption that punishment "errors" are equally costly in either direction (i.e. rewarding an unscrupulous leader and punishing a scrupulous leader both yield the audience a payoff of 0, relative to a payoff of 1 for correctly matching the action to the leader type). This is without loss of generality; we could make the errors asymmetric, inducing a punishment threshold that is different from $\frac{1}{2}$, and all results would follow (as long as the prior belief π remains above a lower bound relative to the threshold, per Assumption 1(iv)).

7.3 Recap of Model Setup and Main Results

To recap, the sequence of moves is as follows:

1. The leader's type $\theta \in \{0,1\}$, with $Pr(\theta = 1) = \pi \in (\frac{1}{2}, 1)$, is realized by Nature and observed privately by the leader.

2. The state variable $\omega \in \{0,1\}$, with $Pr(\omega = 1) = \tau$ and the cost variable k_c, with $F^\theta(x) = Pr(k_c \le x|\theta)$, are realized by Nature and observed privately by the leader. The state variable is independent of the cost and type variables: $\omega \perp\!\!\!\perp \theta$, and $\omega \perp\!\!\!\perp k_c|\theta$.

3. The leader chooses whether to take public action $a_p \in \{0,1\}$, which A observes, and covert action $a_c \in \{0,1\}$, which A does not observe directly.

4. The policy outcome $y \in \{0,1\}$ is realized, according to the probabilities given in (1).

5. The covert revelation $z \in \{0,1\}$ is realized, with $Pr(z = 1|a_c) = a_c\lambda$, where $z \perp\!\!\!\perp y, \omega, \theta, k_c \mid a_c$.

6. The audience observes $(a_p, y, z) \in \{0,1\}^3$, and chooses whether to punish or reward the leader, $r \in \{0,1\}$

7. Payoffs are realized, with $U_A = \mathbb{1}[r = \theta]$ and $U_L = y - a_c k_c - a_p k_p + r\beta$.

Our solution concept is Perfect Bayesian Equilibrium.

Remark 3 (Notation)

- *Let* $r^h = Pr(r = 1|h)$ *denote* A*'s strategy as a function of the history* $h = (a_p, y, z)$
- *Let* $q = r^{010}\beta$, $s = r^{110}\beta$, $t = r^{100}\beta$, $v = r^{000}\beta$, *and let* $\sigma_A = (q, s, t, v) \in [0, \beta]^4$
- *Denote* L*'s action* $a = (a_p, a_c)$
- *Denote the scrupulous leader* L^1 *and the unscrupulous leader* L^0.
- *Denote* $F^{\theta=0}(x)$ *as* $F(x)$.

Assumption 1 (Parameter restrictions)

(i) $\beta < \min\left\{1, 1 + \frac{k_p(\alpha_p^1 - \alpha_{pc}^0)}{\alpha_p^1\alpha_c - \alpha_{pc}^0\alpha_0}\right\}$

(ii) $\alpha_0 < \alpha_c(1 - \lambda)$

(iii) k_p *is in an intermediate range,* $\alpha_p^0 < k_p < \min\left\{\alpha_p^1 - \alpha_0, \alpha_p^1(1 - \alpha_c)\right\}$

(iv) π *is above a lower bound (which requires* $\pi > \frac{1}{2}$*):*

$$\frac{\pi}{(1-\pi)} > \max\left\{1, 1 + F(\alpha_c(1-\alpha_p^1) - \beta\lambda)\left[\frac{\alpha_{pc}^1}{\alpha_p^1}(1-\lambda) - 1\right]\right\}$$

(v) If $k_p < \widehat{k}_p^{NCRE}$ *(as defined in Lemma 8): assume* $\alpha_p^1 \le \bar{\alpha}_p^1$, *where* $\bar{\alpha}_p^1 \in (\alpha_{pc}^0, 1)$ *is the unique threshold such that (10) holds iff* $\alpha_p^1 \le \bar{\alpha}_p^1$.

(vi) k_c *follows a* θ*-specific distribution* $F^\theta(x) = Pr(k_c \le x|\theta)$, *where:*

- $F^{\theta=1}(x) = 0 \, \forall x \leq 1 + \beta$.
- $F = F^{\theta=0}$ *is continuously differentiable with positive density on an open, possibly unbounded, interval* $(\underline{k_c}, \overline{k_c})$*, satisfying*

$$\underline{k_c} < \min_{g \in \mathcal{G}} \min_{\sigma_A \in [0,\beta]^4} g(\sigma_A), \qquad \max_{g \in \mathcal{G}} \max_{\sigma_A \in [0,\beta]^4} g(\sigma_A) < \overline{k_c}$$

for the set of functions $\mathcal{G} = \left\{ k_c^*, \ddot{k}_c^\omega, \overset{\sim\omega}{k_c}, \widehat{k}_c^\omega \right\}_{\omega=0,1}$ *defined in Lemma 1, with these inequalities holding at every* λ *in the range permitted by point (ii), holding the remaining parameters fixed.*

Also recall, as stated in the main text: $0 < \alpha_0 < \alpha_c < 1$; $\alpha_0 \leq \alpha_p^0 < \alpha_p^1 < 1$; and $\lambda, \beta > 0$. Note that the caption of Figure 3 specifies a parameter vector for which all parameter restrictions are jointly satisfied; and since the figure sweeps λ across its full admissible range $(0, \frac{\alpha_c - \alpha_0}{\alpha_c})$, it also demonstrates that A1 holds at every admissible λ while all other parameters are held fixed.

Assumption 1 ensures that the leader faces the strategic problem that motivates us: maintaining plausible deniability for objectionable covert action in the face of strategic inferences.

Point (i) ensures that an equilibrium exists in which the scrupulous leader takes the public action if and only if it is effective (i.e. a Responsive Equilibrium (RE), defined below). Point (ii) means that the probability of success due to exogenous factors is relatively small. When it is violated, the leader's plausible deniability problem becomes trivial. Point (iii) clarifies the distinction between "effective" vs. "ineffective" public action: the lower bound on k_p implies that ineffective public action ($a_p = 1$ when $\omega = 0$) is not worthwhile for policy reasons alone (a point that is central to our concept of a "pure" cover story), while the upper bound implies that effective public action is worthwhile. Point (iv) implies that within an RE, when a cover story is not being used, the leader is not punished for successful public action; this focuses our attention on the problem of "unexplained success", as discussed throughout the main text. Point (v) ensures the existence of a particular form of Cover Story Responsive Equilibrium (CSRE); relaxing this assumption would require demonstrating existence of other, substantively similar CSRE, which we avoid for brevity. Finally, point (vi) implies that the leader's use of covert action varies in response to any marginal change in the audience's punishment/reward strategy (which primarily serves the purpose of avoiding substantively uninteresting edge cases throughout the proofs that follow).

Remark 4 *Covert action is strictly dominated for* L^1.

This is the direct implication of the lower bound on the support of F^1, from Assumption 1 (vi).

Assumption 2 (Off-path beliefs) *At any off-path information set, assume* A *assigns belief* $\mu = \pi$ *if* $z = 0$*, and* $\mu = 0$ *if* $z = 1$.

All results that follow impose Assumptions 1 and 2.

To preview the main results that will be derived throughout the remainder of this appendix (paraphrasing for brevity):

- Proposition 2: An interior Responsive Equilibrium (RE) always exists.
- Proposition 3: There exists a transparency threshold $\bar{\lambda}$ such that: if $\lambda > \bar{\lambda}$, all interior RE are No-Cover Responsive Equilibria (NCRE); and if $\lambda < \bar{\lambda}$, all interior RE are Cover Story Responsive Equilibria (CSRE).

- Proposition 4: There exists a transparency threshold $\bar{\bar{\lambda}} \geq \bar{\lambda}$ such that, in any interior RE, the audience punishes unexplained success iff $\lambda < \bar{\bar{\lambda}}$.

- Proposition 5: If $\lambda > \bar{\lambda}$ and α_0 is small, the interior NCRE exists and yields a higher ex-ante payoff for L^1 than does any other equilibrium.

- Proposition 6: If $\lambda < \bar{\lambda}$ and α_0 is small, L^1's highest ex-ante payoff is achieved by a cover story equilibrium.

7.4 Proofs of Main Text Results, Section 3

7.4.1 Preliminaries

Lemma 1 (L's best response) *Take A's strategy $\sigma_A = (q, s, t, v) \in [0, \beta]^4$ as given (recall the notation from Remark 3), and suppose A plays $r = 0$ whenever $z = 1$. Then L's best response is characterized in Table 2, where $(\widehat{k}_p^{\omega}, \tilde{k}{}_p^{\omega}, \tilde{k}{}_c^{\omega}, \ddot{k}_c^{\omega}, \widehat{k}_c^{\omega}, k_c^*)$ are functions of (q, s, t, v) defined in the proof.*

Table 2: Leader Best Response

		$a = (0,0)$ vs. $a = (1,0)$ thresholds		
		$k_p < \tilde{k}{}_p^{0}$	$\tilde{k}{}_p^{0} < k_p < \tilde{k}{}_p^{1}$	$k_p > \tilde{k}{}_p^{1}$
$a = (0,1)$ vs. $a = (1,1)$ thresholds	$k_p > \widehat{k}_p^1$	ω=0: $(1,0)$ if $k_c > \tilde{k}{}_c^{0}$, else $(0,1)$ ω=1: $(1,0)$ if $k_c > \tilde{k}{}_c^{1}$, else $(0,1)$	ω=0: $(0,0)$ if $k_c > k_c^*$, else $(0,1)$ ω=1: $(1,0)$ if $k_c > \tilde{k}{}_c^{1}$, else $(0,1)$	ω=0: $(0,0)$ if $k_c > k_c^*$, else $(0,1)$ ω=1: $(0,0)$ if $k_c > k_c^*$, else $(0,1)$
	$\widehat{k}_p^0 < k_p < \widehat{k}_p^1$	ω=0: $(1,0)$ if $k_c > \tilde{k}{}_c^{0}$, else $(0,1)$ ω=1: $(1,0)$ if $k_c > \ddot{k}_c^1$, else $(1,1)$	ω=0: $(0,0)$ if $k_c > k_c^*$, else $(0,1)$ ω=1: $(1,0)$ if $k_c > \ddot{k}_c^1$, else $(1,1)$	ω=0: $(0,0)$ if $k_c > k_c^*$, else $(0,1)$ ω=1: $(0,0)$ if $k_c > \widehat{k}_c^1$, else $(1,1)$
	$k_p < \widehat{k}_p^0$	ω=0: $(1,0)$ if $k_c > \ddot{k}_c^0$, else $(1,1)$ ω=1: $(1,0)$ if $k_c > \ddot{k}_c^1$, else $(1,1)$	ω=0: $(0,0)$ if $k_c > \widehat{k}_c^0$, else $(1,1)$ ω=1: $(1,0)$ if $k_c > \ddot{k}_c^1$, else $(1,1)$	ω=0: $(0,0)$ if $k_c > \widehat{k}_c^0$, else $(1,1)$ ω=1: $(0,0)$ if $k_c > \widehat{k}_c^1$, else $(1,1)$

Note: Leader best response $a = (a_p, a_c)$ by public action feasibility ω, public action cost k_p, and covert action cost k_c. Thresholds given in proof of Lemma 1. Recall that the distribution of k_c varies by θ, with $F^{\theta=1}(x) = 0$ for all $x \leq 1 + \beta$. Cutoffs (all from the pairwise comparisons in Lemma 1): $\tilde{k}{}_c^{\omega}$: $(1,0)$ vs. $(0,1)$; $\ddot{k}_c^{\omega}$: $(1,0)$ vs. $(1,1)$; $\widehat{k}_c^{\omega}$: $(0,0)$ vs. $(1,1)$; k_c^*: $(0,0)$ vs. $(0,1)$. Remark 5 will show that $\tilde{k}{}_p^{0} < \tilde{k}{}_p^{1}$ and $\widehat{k}_p^0 < \widehat{k}_p^1$ for any fixed σ_A.

Proof of Lemma 1: All claims in the lemma follow directly from comparing L's expected payoff from each of her available actions. Observe that a strict inequality in a $\widehat{k}_p^{\omega}$ threshold eliminates one $(a_p, a_c = 1)$ action, while a strict inequality in a $\tilde{k}{}_p^{\omega}$ threshold eliminates one $(a_p, a_c = 0)$ action; the two surviving actions are then ranked by whether k_c falls above or below the corresponding k_c threshold, given the realization of ω. (Recall that L^1's covert action cost k_c is always above all k_c thresholds.) The boundaries between each cell in Table 2 denote indifference between $(0, a_c)$ and $(1, a_c)$ for some ω and some k_c interval. Public action indifference in each comparison below holds when k_p equals the corresponding $\tilde{k}{}_p^{\omega}$ or $\widehat{k}_p^{\omega}$ threshold; covert action indifference occurs with zero probability in all cases, per Assumption 1 (vi).

$$
\begin{aligned}
E[U_L(a=(0,0))|\omega,k_c;\sigma_A] &= \alpha_0(1+q)+(1-\alpha_0)v \\
E[U_L(a=(0,1))|\omega,k_c;\sigma_A] &= \alpha_c(1+(1-\lambda)q)+(1-\alpha_c)(1-\lambda)v-k_c \\
E[U_L(a=(1,0))|\omega,k_c;\sigma_A] &= \alpha_p^\omega(1+s)+(1-\alpha_p^\omega)t-k_p \\
E[U_L(a=(1,1))|\omega,k_c;\sigma_A] &= \alpha_{pc}^\omega(1+(1-\lambda)s)+(1-\alpha_{pc}^\omega)(1-\lambda)t-k_p-k_c \\
&\qquad (\text{recall } \alpha_{pc}^\omega=\alpha_p^\omega+\alpha_c(1-\alpha_p^\omega)=\alpha_c+\alpha_p^\omega(1-\alpha_c))
\end{aligned}
$$

$$
E[U_L(a=(0,0))|\omega,k_c;\sigma_A] < E[U_L(a=(1,0))|\omega,k_c;\sigma_A] \iff k_p < \tilde{k}{}_p^\omega
$$
$$
\tilde{k}{}_p^\omega = \alpha_p^\omega-\alpha_0+\alpha_p^\omega s+(1-\alpha_p^\omega)t-\alpha_0 q-(1-\alpha_0)v
$$

$$
E[U_L(a=(0,1))|\omega,k_c;\sigma_A] < E[U_L(a=(1,1))|\omega,k_c;\sigma_A] \iff k_p < \widehat{k}_p^\omega
$$
$$
\widehat{k}_p^\omega = \alpha_p^\omega(1-\alpha_c)+(1-\lambda)[\alpha_{pc}^\omega s+(1-\alpha_{pc}^\omega)t-\alpha_c q-(1-\alpha_c)v]
$$

$$
E[U_L(a=(0,0))|k_c;\sigma_A] < E[U_L(a=(0,1))|k_c;\sigma_A] \iff k_c < k_c^*
$$
$$
k_c^* = \alpha_c-\alpha_0+q[\alpha_c(1-\lambda)-\alpha_0]+v[(1-\alpha_c)(1-\lambda)-(1-\alpha_0)]
$$

$$
E[U_L(a=(0,0))|\omega,k_c;\sigma_A] < E[U_L(a=(1,1))|\omega,k_c;\sigma_A] \iff k_c < \widehat{k}_c^\omega
$$
$$
\widehat{k}_c^\omega = -k_p+\alpha_{pc}^\omega-\alpha_0+(1-\lambda)[\alpha_{pc}^\omega s+(1-\alpha_{pc}^\omega)t]-\alpha_0 q-(1-\alpha_0)v
$$

$$
E[U_L(a=(1,0))|\omega,k_c;\sigma_A] < E[U_L(a=(0,1))|\omega,k_c;\sigma_A] \iff k_c < \tilde{k}{}_c^\omega
$$
$$
\tilde{k}{}_c^\omega = k_p+\alpha_c-\alpha_p^\omega-\alpha_p^\omega s-(1-\alpha_p^\omega)t+\alpha_c(1-\lambda)q+(1-\alpha_c)(1-\lambda)v
$$

$$
E[U_L(a=(1,0))|\omega,k_c;\sigma_A] < E[U_L(a=(1,1))|\omega,k_c;\sigma_A] \iff k_c < \ddot{k}_c^\omega
$$
$$
\ddot{k}_c^\omega = \alpha_c(1-\alpha_p^\omega)+s[\alpha_{pc}^\omega(1-\lambda)-\alpha_p^\omega]+t[(1-\alpha_{pc}^\omega)(1-\lambda)-(1-\alpha_p^\omega)]
$$

Throughout the proofs, the dependence of each threshold on σ_A will generally be suppressed unless it is needed for clarity (e.g. $\tilde{k}{}_p^0(\sigma_A)$ denoted as $\tilde{k}{}_p^0$ for generic σ_A). ■

Remark 5 (Relationship between thresholds) *Considering the thresholds defined in Lemma 1, for any fixed $\sigma_A=(q,s,t,v)$:*

- $\tilde{k}{}_p^\omega - k_p = E[U_L(a=(1,0))|\omega,k_c;\sigma_A]-E[U_L(a=(0,0))|\omega,k_c;\sigma_A] = \widehat{k}_c^\omega-\ddot{k}_c^\omega = k_c^*-\tilde{k}{}_c^\omega$
- $\widehat{k}_p^\omega - k_p = E[U_L(a=(1,1))|\omega,k_c;\sigma_A]-E[U_L(a=(0,1))|\omega,k_c;\sigma_A] = \widehat{k}_c^\omega-k_c^* = \ddot{k}_c^\omega-\tilde{k}{}_c^\omega$
- $\tilde{k}{}_p^1 > \tilde{k}{}_p^0$*, and* $\widehat{k}_p^1 > \widehat{k}_p^0$

For the last point, observe:

- $\tilde{k}{}_p^1 - \tilde{k}{}_p^0 = (\alpha_p^1-\alpha_p^0)(1+s-t) > 0$
- $\widehat{k}_p^1-\widehat{k}_p^0 = (\alpha_p^1-\alpha_p^0)(1-\alpha_c)+(1-\lambda)(\alpha_{pc}^1-\alpha_{pc}^0)(s-t) = (\alpha_p^1-\alpha_p^0)(1-\alpha_c)(1+(1-\lambda)(s-t)) > 0$

Remark 6 (Useful k_p thresholds) *The following bounds will be invoked repeatedly throughout the proofs that follow, to evaluate existence of various equilibria.*

- $\tilde{k}{}_p^0(v=\beta) \leq \alpha_p^0 - \alpha_0 + \alpha_0\beta < \alpha_p^0 < k_p$ *for any* s, t, q
- $\tilde{k}{}_p^1(s=t=\beta) \geq \alpha_p^1 - \alpha_0 > k_p$ *for any* q, v
- $\widehat{k}_p^0(q=v=\beta) \leq \alpha_p^0(1-\alpha_c) < k_p$ *for any* s, t
- $\widehat{k}_p^1(s=t=\beta) \geq \alpha_p^1(1-\alpha_c) > k_p$ *for any* q, v

Note that the inequalities in Remark 5 rely on A1(i) (specifically $\beta < 1$, so $1+s-t \geq 1-\beta > 0$). The inequalities in Remark 6 rely on $\beta < 1$, as well as A1(iii) (upper and lower bounds on k_p).

Remark 7 (Interior probability of covert action) *For any* (σ_A, ω)*, let* $k_c^\dagger \in \left\{k_c^*, \ddot{k}_c^\omega, \tilde{k}{}_c^\omega, \widehat{k}_c^\omega\right\}_{\omega=0,1}$ *denote the covert action threshold governing* L^0*'s best response in the relevant cell of Lemma 1. Under Assumption 1 (vi),* $k_c^\dagger$ *lies strictly within* $\left(\underline{k_c}, \overline{k_c}\right)$*; so in any equilibrium,* $Pr(a_c = 1|\sigma_A, \omega, \theta = 0) = F(k_c^\dagger) \in (0,1)$*. Further, since* F *has positive density at all admissible values of* $k_c^\dagger$*, we have that* F *is strictly increasing on* $\left(\underline{k_c}, \overline{k_c}\right)$*.*

Remark 8 (Positive denominator in Assumption 1 (i)) $\alpha_p^1\alpha_c > \alpha_{pc}^0\alpha_0$.

Proof: Let $D := \alpha_p^1\alpha_c - \alpha_0\alpha_{pc}^0$. To show that $D > 0$: rewrite $\alpha_{pc}^0 = \alpha_p^0 + \alpha_c - \alpha_p^0\alpha_c$, and rearrange to

$$D = \alpha_c(\alpha_p^1 - \alpha_0) - \alpha_0\alpha_p^0(1-\alpha_c) > \alpha_c\alpha_p^0 - \alpha_0\alpha_p^0(1-\alpha_c) = \alpha_p^0(\alpha_c - \alpha_0(1-\alpha_c)) > 0$$

The first inequality follows from the fact that $\alpha_p^1 - \alpha_0 > \alpha_p^0$, per Assumption 1 (iii). The last inequality follows from the assumption that $\alpha_c > \alpha_0$. ■

Note that D is the denominator in the expression in Assumption 1 (i). (This will be referenced in the proofs of Lemma 9 and Proposition 6.)

Lemma 2 (Threshold strategy in audience beliefs) *In every equilibrium, given on-path history* $h = (a_p, y, z)$*:*

- *The audience's best response satisfies*

$$r = \begin{cases} 0, & \mu^h < \frac{1}{2} \\ 1, & \mu^h > \frac{1}{2} \\ \text{any } r \in [0,1], & \mu^h = \frac{1}{2} \end{cases}, \quad \text{where } \mu^h = Pr(\theta = 1|h) \tag{7}$$

- *If covert action is directly revealed* ($z = 1$)*, the audience fully punishes the leader* ($r = 0$)*.*

Proof of Lemma 2: Equation (7) follows directly from the audience's utility function, given in (3). The second point of the lemma follows from the fact that L^1 never takes covert action (Remark 4), while L^0 does so with interior probability (Remark 7); so any on-path $z = 1$ history induces beliefs $\mu^h = 0$ by Bayes' Rule. ■

Table 3: History probabilities $P_\theta(a_p, y)$ shaping audience beliefs

$\theta = 1$ contributions: $P_1(a_p, y) = Pr(a_p, y, z = 0|\theta = 1)$

$k_p < \tilde{k}_p^0$	$\tilde{k}_p^0 < k_p < \tilde{k}_p^1$	$k_p > \tilde{k}_p^1$
$P_1(1,y){=}\tau g_y(\alpha_p^1) + (1{-}\tau)g_y(\alpha_p^0)$ $P_1(0,y){=}0$	$P_1(1,y){=}\tau g_y(\alpha_p^1)$ $P_1(0,y){=}(1{-}\tau)g_y(\alpha_0)$	$P_1(1,y){=}0$ $P_1(0,y){=}g_y(\alpha_0)$

$\theta = 0$ contributions: $P_0(a_p, y) = Pr(a_p, y, z = 0|\theta = 0)$

	$k_p < \tilde{k}_p^0$	$\tilde{k}_p^0 < k_p < \tilde{k}_p^1$	$k_p > \tilde{k}_p^1$
$k_p > \hat{k}_p^1$	$P_0(1,y){=}\tau(1{-}F(\tilde{k}_c^1))g_y(\alpha_p^1)$ $+(1{-}\tau)(1{-}F(\tilde{k}_c^0))g_y(\alpha_p^0)$ $P_0(0,y){=}g_y(\alpha_c)(1{-}\lambda)\times$ $[\tau F(\tilde{k}_c^1) + (1{-}\tau)F(\tilde{k}_c^0)]$	$P_0(1,y){=}\tau(1{-}F(\tilde{k}_c^1))g_y(\alpha_p^1)$ $P_0(0,y){=}\tau F(\tilde{k}_c^1)g_y(\alpha_c)(1{-}\lambda)$ $+(1{-}\tau)F(k_c^*)g_y(\alpha_c)(1{-}\lambda)$ $+(1{-}\tau)(1{-}F(k_c^*))g_y(\alpha_0)$	$P_0(1,y){=}0$ $P_0(0,y){=}F(k_c^*)g_y(\alpha_c)(1{-}\lambda)$ $+(1{-}F(k_c^*))g_y(\alpha_0)$
$\hat{k}_p^0 < k_p < \hat{k}_p^1$	$P_0(1,y){=}\tau[(1{-}F(\ddot{k}_c^1))g_y(\alpha_p^1)$ $+F(\ddot{k}_c^1)g_y(\alpha_{pc}^1)(1{-}\lambda)]$ $+(1{-}\tau)(1{-}F(\tilde{k}_c^0))g_y(\alpha_p^0)$ $P_0(0,y){=}g_y(\alpha_c)(1{-}\lambda)(1{-}\tau)F(\tilde{k}_c^0)$	$P_0(1,y){=}\tau[(1{-}F(\ddot{k}_c^1))g_y(\alpha_p^1)$ $+F(\ddot{k}_c^1)g_y(\alpha_{pc}^1)(1{-}\lambda)]$ $P_0(0,y){=}(1{-}\tau)[F(k_c^*)g_y(\alpha_c)(1{-}\lambda)$ $+(1{-}F(k_c^*))g_y(\alpha_0)]$	$P_0(1,y){=}\tau F(\hat{k}_c^1)g_y(\alpha_{pc}^1)(1{-}\lambda)$ $P_0(0,y){=}(1{-}\tau)[F(k_c^*)g_y(\alpha_c)(1{-}\lambda)$ $+(1{-}F(k_c^*))g_y(\alpha_0)]$ $+\tau(1{-}F(\hat{k}_c^1))g_y(\alpha_0)$
$k_p < \hat{k}_p^0$	$P_0(1,y){=}\tau F(\ddot{k}_c^1)g_y(\alpha_{pc}^1)(1{-}\lambda)$ $+(1{-}\tau)F(\ddot{k}_c^0)g_y(\alpha_{pc}^0)(1{-}\lambda)$ $+\tau(1{-}F(\ddot{k}_c^1))g_y(\alpha_p^1)$ $+(1{-}\tau)(1{-}F(\ddot{k}_c^0))g_y(\alpha_p^0)$ $P_0(0,y){=}0$	$P_0(1,y){=}\tau[(1{-}F(\ddot{k}_c^1))g_y(\alpha_p^1)$ $+F(\ddot{k}_c^1)g_y(\alpha_{pc}^1)(1{-}\lambda)]$ $+(1{-}\tau)F(\hat{k}_c^0)g_y(\alpha_{pc}^0)(1{-}\lambda)$ $P_0(0,y){=}(1{-}\tau)(1{-}F(\hat{k}_c^0))g_y(\alpha_0)$	$P_0(1,y){=}\tau F(\hat{k}_c^1)g_y(\alpha_{pc}^1)(1{-}\lambda)$ $+(1{-}\tau)F(\hat{k}_c^0)g_y(\alpha_{pc}^0)(1{-}\lambda)$ $P_0(0,y){=}(1{-}\tau)(1{-}F(\hat{k}_c^0))g_y(\alpha_0)$ $+\tau(1{-}F(\hat{k}_c^1))g_y(\alpha_0)$

Note: $g_y(\alpha) = y\alpha + (1-y)(1-\alpha)$. Each entry denotes the probability of leader type θ reaching history $h = (a_p, y; z = 0)$, given their best-response from Lemma 1, represented by where k_p falls relative to each threshold $\tilde{k}_p^\omega, \hat{k}_p^\omega$. Each $F(\cdot)$ threshold is the covert cutoff governing L^0's strategy in the corresponding cell of Table 2. L^1's strategy does not depend on $\hat{k}_p^\omega$ thresholds.

Lemma 3 (Generic form of on-path audience beliefs) *Let $h = (a_p, y)$ denote a history of the game with $z = 0$. By Bayes' Rule, in any on-path history h, the audience's beliefs satisfy*

$$\mu^h = Pr(\theta = 1|h) = \frac{\pi P_1(h)}{\pi P_1(h) + (1-\pi)P_0(h)}$$

where $P_\theta(h) = Pr(h|\theta)$. Then, $\mu^h \geq \frac{1}{2} \iff \frac{\pi}{1-\pi}P_1(h) \geq P_0(h)$, and $\mu^h \geq \pi \iff P_1(h) \geq P_0(h)$. For any h, h' such that $P_1(h), P_1(h') > 0$, $\mu^h > \mu^{h'} \iff \frac{P_0(h)}{P_1(h)} < \frac{P_0(h')}{P_1(h')}$.

Lemma 4 (Necessary condition for cover story) *L^0 uses a cover story with positive probability only if $k_p \leq \hat{k}_p^0$.*

Proof of Lemma 4: A cover story requires playing $a = (1, 1)$ when $\omega = 0$. If $k_p > \hat{k}_p^0$, then L^0 strictly prefers $a = (0, 1)$ over $a = (1, 1)$ when $\omega = 0$, per Lemma 1. ■

7.4.2 Within-equilibrium analysis: the Responsive Equilibrium (RE)

This section characterizes behavior within a substantively appealing class of equilibria, which we define as (interior) Responsive Equilibria (RE). Proposition 2 shows that this class of equilibrium always exists. The next two propositions partition interior RE behavior into three regions, as cutpoints in the transparency parameter λ:

- Proposition 4 says that the audience punishes unexplained success (i.e. plays $q < \beta$) iff $\lambda < \bar{\bar{\lambda}}$, where $\bar{\bar{\lambda}} \in \left[\bar{\lambda}, \frac{\alpha_c - \alpha_0}{\alpha_c}\right)$.
- Proposition 3 says that the leader plays a cover story with positive probability if $\lambda < \bar{\lambda}$, and does not play a cover story if $\lambda > \bar{\lambda}$.

The propositions build from the following intermediate results:

- Lemma 7 provides a general characterization of audience strategies following public inaction in any interior RE.
- Lemma 8 demonstrates existence of a No-Cover Responsive Equilibrium (NCRE) when $\lambda \geq \bar{\lambda}$ (equivalently, when $k_p \geq \widehat{k}_p^{NCRE}$), and uniqueness when the inequality is strict.
- Lemma 9 demonstrates existence of a Cover Story Responsive Equilibrium (CSRE) when $\lambda < \bar{\lambda}$ (equivalently, when $k_p < \widehat{k}_p^{NCRE}$).
- Lemma 10 establishes the correspondence between the $\bar{\lambda}$ and $\widehat{k}_p^{NCRE}$ thresholds.
- Lemmas 8 and 9 jointly establish the universal existence of interior RE (Proposition 2); and that among interior RE, a threshold $\bar{\lambda}$ determines whether the equilibrium is a NCRE or a CSRE (Proposition 3).
- Proposition 4 is a comparative static result within the interior NCRE, drawing on the audience best-responses established in Lemmas 7, 8, and 9.

Finally, a series of corollaries provide additional insight into the strategic logic of the interior CSRE:

- Corollary 1 shows how the $\bar{\lambda}$ threshold for the use of cover stories varies as a function of parameters α_c, k_p, β.
- Corollaries 2 and 3 highlight how the audience's interim beliefs operate within a CSRE.
- Corollary 4 shows cover stories can only be used with interior probability in a CSRE.

Definition 3 *Define a* ***Responsive Equilibrium (RE)*** *as an equilibrium in which L^1 plays $a_p = \omega$. Define an* ***interior RE*** *as an RE satisfying $\overset{\sim 0}{k_p} < k_p < \overset{\sim 1}{k_p}$.*

Remark 9 ($\overset{\sim\omega}{\boldsymbol{k_p}}$ bounds for RE) *An RE requires $\overset{\sim 0}{k_p} \leq k_p \leq \overset{\sim 1}{k_p}$ (where $\overset{\sim\omega}{k_p}(\sigma_A)$ are endogenous functions of the audience strategy σ_A). Any equilibrium in which $\overset{\sim 0}{k_p} < k_p < \overset{\sim 1}{k_p}$ is an RE.*

Proof of Remark 9: Recall that an RE is defined as an equilibrium in which L^1 plays $a_p = \omega$, and recall that L^1 never plays $a_c = 1$ (given that $\inf \operatorname{supp}(F^1) \geq 1 + \beta$). The thresholds $\overset{\sim 0}{k_p}, \overset{\sim 1}{k_p}$ relative to k_p determine L^1's preference for $a = (0,0)$ vs. $a = (1,0)$. An RE requires that L^1 weakly prefers $a_p = \omega$; and if k_p is strictly bounded within the thresholds, then the RE strategy is L^1's unique best response. Note that L^0's strategy (both the a_c and a_p components) may differ from L^1's within an RE, as we will see below. ■

Lemma 5 *In any interior RE, $k_p \geq \widehat{k}_p^0$.*

Proof of Lemma 5: Suppose we have an interior RE with $k_p < \widehat{k}_p^0$. In this equilibrium, following the notation of Lemma 3 (and suppressing $z = 0$ from the $h = (a_p, y)$ notation throughout, for succinctness), we have

$$P_1(0,1) = (1-\tau)\alpha_0 \geq P_0(0,1) = (1-\tau)\alpha_0[1 - F(\widehat{k}_c^0(\cdot))]$$
$$P_1(0,0) = (1-\tau)(1-\alpha_0) \geq P_0(0,0) = (1-\tau)(1-\alpha_0)[1 - F(\widehat{k}_c^0(\cdot))]$$

so $\mu^{000}, \mu^{010} \geq \pi$ and thus $q = v = \beta$. But $\widehat{k}_p^0(v = q = \beta) < k_p$ (Remark 6), contradicting $\widehat{k}_p^0 > k_p$. ■

Lemma 6 (Leader strategies in an interior RE) *Consider an interior RE satisfying $k_p < \widehat{k}_p^1$. In any such equilibrium, the leader's strategies are characterized as follows:*[20]

- *L^1 plays $a_p = \omega$ and $a_c = 0$.*
- *When $\omega = 1$: L^0 plays $a_p = 1$, and $a_c = \mathbb{1}[k_c < \ddot{k}_c^1]$.*
- *When $\omega = 0$:*
  - *if $k_c > k_c^*$, L^0 plays $a = (0, 0)$;*
  - *if $k_c < k_c^*$ and $k_p > \widehat{k}_p^0$, L^0 plays $a = (0, 1)$;*
  - *if $k_c < k_c^*$ and $k_p = \widehat{k}_p^0$, L^0 plays a (possibly degenerate) mixture over $a = (0, 1)$ and $a = (1, 1)$.*

Proof of Lemma 6: L^1's strategy follows from the definition of RE, and from Remark 4. For L^0's strategy, observe that the stipulated conditions (along with Lemma 5) place us in the cell of Table 2 with $\widehat{k}_p^0 \leq k_p < \widehat{k}_p^1$, $\tilde{k}{}_p^0 < k_p < \tilde{k}{}_p^1$. ■

Lemma 7 (Audience BR to public inaction in interior RE) *Let*

$$\mu^{010}(q, \phi) := \frac{\pi\alpha_0}{\pi\alpha_0 + (1-\pi)[\alpha_0 + F(k_c^*(q, v = \beta))((1-\phi)\alpha_c(1-\lambda) - \alpha_0)]}$$

where $\phi = Pr(a_p = 1|\omega = 0, a_c = 1)$.[21] *Let $\bar{\phi} := 1 - \frac{\alpha_0}{\alpha_c(1-\lambda)}$. Let*

$$q^*(\phi) := \begin{cases} 0, & \mu^{010}(0,\phi) < \frac{1}{2} \\ \widehat{q}, & \mu^{010}(\beta,\phi) \leq \frac{1}{2} \leq \mu^{010}(0,\phi) \\ \beta, & \mu^{010}(\beta,\phi) > \frac{1}{2} \end{cases}, \qquad \text{where } \widehat{q} \text{ uniquely solves } \mu^{010}(\widehat{q},\phi) = \frac{1}{2}$$

Then, in any interior RE characterized by $k_p < \widehat{k}_p^1$, we have the following:

[20] At measure-zero sets $k_c = k_c^*$ and $k_c = \ddot{k}_c^1$, L^0 is free to break ties arbitrarily; this does not affect payoffs or audience beliefs, so we suppress such qualifications throughout the proofs.

[21] ϕ denotes an aggregate conditional probability $Pr(a_p = 1|\omega = 0, a_c = 1)$, integrating over all k_c for which L^0 plays $a_c = 1$ when $\omega = 0$. Regardless of whether different k_c-types apply different mixing probabilities (conditional on $\omega = 0, a_c = 1$), the audience's posterior beliefs only depend on the aggregate probability ϕ.

- *If $\phi \geq \bar{\phi}$, then $\mu^{010} \geq \pi$ and $q^*(\phi) = \beta$.*
- *The audience's beliefs satisfy $\mu^{000} > \frac{1}{2}$, and their strategy satisfies $v = \beta$.*
- *The audience's beliefs satisfy $\mu^{010} = \mu^{010}(q^*(\phi), \phi)$, and their strategy satisfies $q = q^*(\phi)$.*
- *If $\phi < \bar{\phi}$, then $q^*(\phi)$ is continuous, weakly increasing in ϕ, and strictly increasing for $q^*(\phi) \in (0, \beta)$.*
- *There exists a $\lambda' < \frac{\alpha_c - \alpha_0}{\alpha_c}$ such that $q^*(\phi) = \beta$ for every $\lambda \in \left(\lambda', \frac{\alpha_c - \alpha_0}{\alpha_c}\right)$ and every $\phi \in [0, 1]$.*

Proof of Lemma 7: The displayed expression for $\mu^{010}(q, \phi)$ is a direct application of Bayes' Rule, given the leader strategies from Lemma 6 and given $v = \beta$. Then, take each point in turn:

First, $\phi \geq \bar{\phi}$ implies that the quantity in square brackets in the denominator of the $\mu^{010}(q, \phi)$ expression is $\leq \alpha_0$, and thus $\mu^{010} \geq \pi$ for any q, v. So $q = \beta$ is the audience's unique best-response whenever $\phi \geq \bar{\phi}$.

Second, following the notation from Lemma 3, audience beliefs in any RE satisfy

$$P_1(0,0) = (1-\tau)(1-\alpha_0), \qquad P_0(0,0) = (1-\tau)[\kappa_0(1-\alpha_c)(1-\lambda)(1-\phi) + (1-\kappa_0)(1-\alpha_0)]$$

where $\kappa_0 = F(k_c^*(q, v))$. Then, because $(1-\alpha_c) < (1-\alpha_0)$, it follows that for any $\lambda, \phi, \kappa_0 \in [0, 1]$, we have $P_0(0,0) \leq P_1(0,0)$, which implies $\mu^{000} \geq \pi$ and thus $v = \beta$.

Third, to give an intuition for the fixed-point logic: when q increases, L^0's incentive to use covert action also increases, which in turn worsens A's belief in the (0,1,0) history and causes them to decrease q. The fixed point $q^*(\phi)$ is then the audience's best-response to the leader behavior induced by the audience's reward strategy. More precisely: A's belief μ^{010} follows from the expressions $P_1(0,1)$, $P_0(0,1)$, analogously to the previous point (replacing $(1-\alpha_0), (1-\alpha_c)$ with α_0, α_c). Fix $\phi < \bar{\phi}$. (The $\phi \geq \bar{\phi}$ case was handled in point 1.) We know $\frac{\partial k_c^*}{\partial q} = \alpha_c(1-\lambda) - \alpha_0 > 0$ (A1(ii)); and F has positive density at $k_c^*(q, v = \beta)$ (A1(vi)); and μ^{010} is decreasing in $F(k_c^*)$. So the map $q \mapsto \mu^{010}(q, \phi)$ is continuous and strictly decreasing. If $\mu^{010}(q, \phi) < \frac{1}{2}$ holds for $q = 0$, then it holds for any q, so A's best-response dictates $q = 0$. Likewise, if $\mu^{010}(q, \phi) > \frac{1}{2}$ holds for $q = \beta$, then it holds for any q, so A's best-response dictates $q = \beta$. In the interior branch, $F(k_c^*(q, v = \beta))$ strictly increases in q. Thus there exists a unique fixed point $\widehat{q}$ that solves $\mu^{010}(\widehat{q}, \phi) = \frac{1}{2}$. In all three cases, the unique belief-consistent value is $q = q^*(\phi)$.

Fourth, continuity of $q^*(\phi)$ follows from the previous point, and the fact that the endpoints of the interior branch ($q^*(\phi) = \hat{q}$) equal 0 when $\mu^{010}(0, \phi) = \frac{1}{2}$, and β when $\mu^{010}(\beta, \phi) = \frac{1}{2}$. Monotonicity of $q^*(\phi)$ follows from implicit differentiation, $\frac{\partial q^*(\phi)}{\partial \phi} = -\frac{\partial \mu^{010}/\partial \phi}{\partial \mu^{010}/\partial q}$: the numerator is positive; the denominator is negative for $q^*(\phi) \in (0, \beta)$ (with $q^*(\phi)$ flat in ϕ at the edges, when $q^*(\phi) \in \{0, \beta\}$).

Finally, we want to show that $\mu^{010}(\beta, \phi) > \frac{1}{2}$ for all λ sufficiently close to $\frac{\alpha_c - \alpha_0}{\alpha_c}$, and all $\phi \in [0, 1]$. Since $\mu^{010}(\beta, \phi)$ is increasing in ϕ, it suffices to establish the claim at $\phi = 0$. As $\lambda \to \frac{\alpha_c - \alpha_0}{\alpha_c}$, we have $\alpha_c(1-\lambda) - \alpha_0 \to 0$, so $\mu^{010}(\beta, 0) \to \frac{\pi\alpha_0}{\pi\alpha_0 + (1-\pi)\alpha_0} = \pi > \frac{1}{2}$. By continuity of $\mu^{010}(\beta, 0)$ in λ, there exists $\lambda' < \frac{\alpha_c - \alpha_0}{\alpha_c}$ such that $\mu^{010}(\beta, 0) > \frac{1}{2}$—and hence $\mu^{010}(\beta, \phi) > \frac{1}{2}$ for every $\phi \in [0, 1]$—for all $\lambda \in \left(\lambda', \frac{\alpha_c - \alpha_0}{\alpha_c}\right)$. ■

Lemma 8 (NCRE existence and uniqueness) *Following the notation of ϕ, $q^*(\phi)$ from Lemma 7: Define a **No-Cover RE (NCRE)** as an RE characterized by $\phi = 0$. Denote*

$$\widehat{k}_p^{NCRE} := \widehat{k}_p^0(s = t = v = \beta, q = q^*(\phi = 0)), \text{ and observe } \widehat{k}_p^{NCRE} = \alpha_p^0(1-\alpha_c) + (1-\lambda)\alpha_c(\beta - q^*(0))$$

Then:

- *If $\widehat{k}_p^{NCRE} < k_p$, a unique interior RE exists, and it is an NCRE that satisfies $k_p < \widehat{k}_p^1$.*
- *An interior NCRE exists if and only if $\widehat{k}_p^{NCRE} \leq k_p$.*

Proof of Lemma 8: To outline the proof, we will show the following:

- Any interior RE satisfying $k_p > \widehat{k}_p^{NCRE}$ must feature $\phi = 0$.
- Suppose an interior RE with $\phi = 0$. These conditions pin down a unique equilibrium;[22] this equilibrium satisfies $k_p < \widehat{k}_p^1$, and it exists if and only if $k_p \geq \widehat{k}_p^{NCRE}$.

Consider an equilibrium satisfying $\tilde{k}_p^{0} < k_p < \tilde{k}_p^{1}$. From Remark 9 (and Definition 3), this must be an interior RE; and thus from Lemma 5, it must satisfy $k_p \geq \widehat{k}_p^0$.

Suppose the equilibrium satisfies $\widehat{k}_p^{NCRE} < k_p$, and $\phi = \phi'$ for some $\phi' > 0$. Then it must be that $k_p = \widehat{k}_p^0$ (by Lemmas 4 and 5), and thus $k_p < \widehat{k}_p^1$ (Remark 5). From Lemma 7, it follows that $v = \beta$ and $q = q^*(\phi') \geq q^*(0)$ (as q^* is weakly increasing in ϕ). Note that $\widehat{k}_p^0$ is increasing in s, t and decreasing in q, v. So $k_p = \widehat{k}_p^0(s, t, q = q^*(\phi'), v = \beta) \leq \widehat{k}_p^{NCRE} = \widehat{k}_p^0(s = t = v = \beta, q = q^*(0))$; but this contradicts $k_p > \widehat{k}_p^{NCRE}$. Thus we cannot have an interior RE with $k_p > \widehat{k}_p^{NCRE}$ and $\phi > 0$.

Now suppose the interior RE satisfies $\phi = 0$. The next step is to determine audience beliefs at μ^{1y0} and best responses s, t. From Table 3, in the cells satisfying $\widehat{k}_p^0 \leq k_p$, $\tilde{k}_p^{0} < k_p < \tilde{k}_p^{1}$, we have:

$$P_1(1,1) = \tau\alpha_p^1, \quad P_1(1,0) = \tau(1-\alpha_p^1), \quad P_1(0,1) = (1-\tau)\alpha_0, \quad P_1(0,0) = (1-\tau)(1-\alpha_0)$$

$$P_0(1,y) = m_{pc}^1 g_y(\alpha_{pc}^1) + m_p^1 g_y(\alpha_p^1), \qquad g_y(\alpha) = y\alpha + (1-y)(1-\alpha)$$

$$m_{pc}^1 = \tau\psi\kappa_1(1-\lambda), \quad m_p^1 = \tau(1-\kappa_1), \qquad \kappa_1 = \begin{cases} F(\tilde{k}_c^{1}(q,s,t,v)) & k_p > \widehat{k}_p^1 \\ F(\ddot{k}_c^1(s,t)) & otherwise \end{cases}$$

$$\psi = Pr(a = (1,1)|\theta = 0, \omega = 1, k_c < \{\ddot{k}_c^1 \text{ or } \tilde{k}_c^{1}\})$$

Note that ψ can be properly mixed only at $k_p = \widehat{k}_p^1$; otherwise $\psi = \mathbb{1}[k_p < \widehat{k}_p^1]$.

$m_{pc}^1 + m_p^1 = \tau\left[\kappa_1\psi(1-\lambda) + (1-\kappa_1)\right] \leq \tau$, and $(1-\alpha_{pc}^1) < (1-\alpha_p^1)$, so $P_0(1,0) \leq P_1(1,0)$ for all $\kappa_1, \psi \in [0,1]$. Therefore $\mu^{100} \geq \pi$, so $t = \beta$ is the audience's unique best response.

We next want to show that the equilibrium requires $\mu^{110} > \frac{1}{2}$. To evaluate μ^{110}, first suppose $k_p > \widehat{k}_p^1$. Then $\psi = 0$ and $m_{pc}^1 = 0$; so we can see that $P_0(1,1) \leq P_1(1,1)$, which implies $\mu^{110} \geq \pi$. Thus the audience's unique best-response is $s = \beta$; but $\widehat{k}_p^1(s = t = \beta) > k_p$ for all q, v (Remark 6), contradicting $k_p > \widehat{k}_p^1$. So we know $k_p \leq \widehat{k}_p^1$, so the governing covert-action threshold is $\ddot{k}_c^1$.

Note that $P_0(1,1)$ is increasing in ψ. Plugging in $\psi = 1$ to get an upper bound on $P_0(1,1)$, we have that $\mu^{110} > \frac{1}{2}$ if

$$\frac{\pi}{1-\pi} > 1 + \frac{1}{\alpha_p^1}\kappa_1\left[\alpha_{pc}^1(1-\lambda) - \alpha_p^1\right] \tag{8}$$

If $\alpha_{pc}^1(1-\lambda) \leq \alpha_p^1$, the inequality is satisfied given $\frac{\pi}{1-\pi} > 1$ (per Assumption 1 (iv)). Otherwise, differentiating the RHS of (8) w.r.t. s (while holding fixed $t = \beta$) gives

$$\frac{1}{\alpha_p^1}\left(\alpha_{pc}^1(1-\lambda) - \alpha_p^1\right)F'(\ddot{k}_c^1)\frac{\partial \ddot{k}_c^1}{\partial s} = \frac{1}{\alpha_p^1}\left(\alpha_{pc}^1(1-\lambda) - \alpha_p^1\right)^2 F'(\ddot{k}_c^1) \geq 0$$

[22]Unique, up to measure-zero sets of unscrupulous types with $k_c \in \left\{k_c^*, \ddot{k}_c^1, \tilde{k}_c^{1}\right\}$ who may be indifferent between using covert action and not.

which means that the RHS of (8) is maximized at $s = \beta$. The second option for the lower bound on π in Assumption 1 (iv) is equivalent to (8) with $s = t = \beta$ plugged in to the κ_1 term. Therefore we know that (8) is satisfied for any $s \in [0, \beta]$. As such, $s = \beta$ is the audience's unique best response.

From $s = t = \beta$, we can see that $\widehat{k}_p^1(s = t = \beta) > k_p$ and $\tilde{k}{}_p^1(s = t = \beta) > k_p$ for all q, v (Remark 6). Then Lemma 7 tells us that the audience's unique best response features $v = \beta$ and $q = q^*(\phi = 0)$, which in turn satisfies $\tilde{k}{}_p^0(v = \beta) < k_p$. Finally, $\widehat{k}_p^0(s = t = v = \beta, q = q^*(0)) = \widehat{k}_p^{NCRE}$. So if $\widehat{k}_p^{NCRE} \leq k_p$, all conditions for NCRE existence are satisfied; but if $\widehat{k}_p^{NCRE} > k_p$, the NCRE is not supported.

To summarize: If $k_p > \widehat{k}_p^{NCRE}$, an interior NCRE exists satisfying $k_p < \widehat{k}_p^1$, and it is the unique interior RE that exists. If $k_p = \widehat{k}_p^{NCRE}$ (a measure-zero set), an interior NCRE exists (but may not be unique, among interior RE). If $k_p < \widehat{k}_p^{NCRE}$, an interior NCRE does not exist. ■

Lemma 9 (CSRE existence) *Following the notation of ϕ, $\bar{\phi}$, $q^*(\phi)$, and $\widehat{k}_p^{NCRE}$ from Lemmas 7 and 8: Define a* ***Cover Story Responsive Equilibrium (CSRE)*** *as an RE characterized by $\phi > 0$. If $\widehat{k}_p^{NCRE} > k_p$, an interior CSRE exists (that is, a CSRE satisfying $\tilde{k}{}_p^0 < k_p < \tilde{k}{}_p^1$).*

Proof of Lemma 9: Given $\widehat{k}_p^{NCRE} > k_p$, we will demonstrate the existence of a CSRE characterized by $\phi = \phi^* > 0$ and $\widehat{k}_p^0(q^*(\phi^*), s^*(\phi^*), t = \beta, v = \beta) = k_p$, with ϕ^* and $s^*(\phi)$ defined below. The proof will proceed through the following steps:

- Define $s^*(\phi)$ as the s satisfying $\widehat{k}_p^0(q = q^*(\phi), s^*(\phi), t = \beta, v = \beta) = k_p$.
- Posit an audience strategy $\sigma_A^*(\phi) = (q = q^*(\phi), s = s^*(\phi), t = \beta, v = \beta)$, and a best-response for the leader. Verify that the leader's best-response implies $\mu^{110} \leq \mu^{100}$.
- Show that the leader's best response to $\sigma_A^*(\phi)$ induces audience belief $\mu^{110}(\phi)$, continuous and strictly decreasing on $[0, \hat{\phi}]$ with $\mu^{110}(0) > \frac{1}{2}$. This pins down a $\phi^* \in (0, \hat{\phi}]$ at which $\mu^{110}(\phi^*) \geq \frac{1}{2}$ and $s = s^*(\phi^*)$ is an audience best-response. Together with $\mu^{110} \leq \mu^{100}$, it follows that $t = \beta$ is an audience best-response.
- Verify that $\tilde{k}{}_p^1(\sigma_A^*(\phi^*)) > k_p$.

Defining terms: Consider a strategy profile $\sigma_A^*(\phi) = (q = q^*(\phi), s = s^*(\phi), t = v = \beta)$ for some $\phi > 0$, where $q^*(\phi)$ comes from Lemma 7 and $s = s^*(\phi)$ is the unique solution to $\widehat{k}_p^0(s, t = v = \beta, q = q^*(\phi)) = k_p$:

$$s^*(\phi) = \frac{1}{\alpha_{pc}^0}\left[\frac{k_p - \alpha_p^0(1-\alpha_c)}{1-\lambda} + \alpha_p^0(1-\alpha_c)\beta + \alpha_c q^*(\phi)\right]$$

Let $\phi^{q=0}$ denote $\max\{\phi : q^*(\phi) = 0\}$ if it exists (and $\phi^{q=0} = 0$ otherwise), and $\phi^{q=\beta}$ denote $\min\{\phi : q^*(\phi) = \beta\}$ (which always exists, given that $q^*(\phi \geq \bar{\phi}) = \beta$). We can see that $s^*(\phi)$ is continuous and strictly increasing on $[\phi^{q=0}, \phi^{q=\beta}]$ (inheriting these properties from the interior branch of $q^*(\phi)$ in Lemma 7); and further, that $s^*(\phi^{q=0}) < \beta < s^*(\phi^{q=\beta})$. (The first inequality follows from $\widehat{k}_p^{NCRE} > k_p$: at $q = 0 \leq q^*(0)$, an $s^* < \beta$ is needed to bring $\widehat{k}_p^0$ down to k_p. The second follows from the fact that $\widehat{k}_p^0(q = v = t = \beta) < k_p$ (Remark 6), so an $s^* > \beta$ would be needed to bring it up to k_p.) Thus there exists a unique $\hat{\phi} \in (\phi^{q=0}, \phi^{q=\beta})$ that solves $s^*(\hat{\phi}) = \beta$.

Let $q^\sharp := q^*(\hat{\phi})$; observe that $q^\sharp = \beta - \left(\dfrac{k_p - \alpha_p^0(1-\alpha_c)}{(1-\lambda)\alpha_c}\right) \in (0, \beta)$. Let $s_0 = s^*(0)$; note that $s_0 > 0$, and that $s^*(\phi) = s_0$ for all $\phi \leq \phi^{q=0}$.

Leader strategy: We posit that the leader strategy characterized in Lemma 6, with cover story probability $\phi \in (0,1)$, is a best-response to the stipulated audience strategy $\sigma_A^*(\phi)$. The incentive-compatibility conditions needed to support this leader strategy are as follows:

- L^1's strategy is supported if $\tilde{k}{}_p^0 \le k_p \le \tilde{k}{}_p^1$:
  - Given that the posited audience strategy features $v = \beta$, we have $\tilde{k}{}_p^0(v = \beta) < k_p$ for any s, t, q (Remark 6).
  - $k_p < \tilde{k}{}_p^1$ will be verified below (the strict inequality being necessary for this equilibrium to be an *interior* CSRE).
- L^0's covert action best-response is governed by the thresholds $k_c^*, \ddot{k}_c^1$, per Lemma 1.
- Given $\widehat{k}_p^0 = k_p$ (which holds under $\sigma_A^*(\phi)$, by definition), L^0's indifference condition is satisfied, enabling her to play any $\phi \in [0,1]$ when $\omega = 0, k_c \le k_c^*$.
- Since $\widehat{k}_p^0 < \widehat{k}_p^1$ (by Remark 5), $\widehat{k}_p^0 = k_p \implies k_p < \widehat{k}_p^1$.

Verifying $\mu^{110} \le \mu^{100}$: For any s, t, ϕ, audience beliefs are given by Lemma 3, with

$$P_1(1,y) = \tau g_y(\alpha_p^1), \qquad g_y(\alpha) = y\alpha + (1-y)(1-\alpha)$$
$$P_0(1,y) = m_{pc}^1(\cdot) g_y(\alpha_{pc}^1) + m_p^1(\cdot) g_y(\alpha_p^1) + m_{pc}^0(\cdot) g_y(\alpha_{pc}^0)$$

$$m_{pc}^1(s,t) = \tau(1-\lambda)\kappa_1(s,t), \quad m_p^1(s,t) = \tau(1-\kappa_1(s,t)), \quad m_{pc}^0(q,\phi) = (1-\tau)\phi(1-\lambda)\kappa_0(\phi)$$
$$\kappa_1(s,t) = F(\ddot{k}_c^1(s,t)), \quad \kappa_0(\phi) = F(k_c^*(q = q^*(\phi), v = \beta))$$

Now consider a CSRE featuring $\sigma_A^*(\phi)$ and $\phi \in [0, \hat{\phi}]$. In any such equilibrium, $\mu^{110} \le \mu^{100}$ is equivalent to

$$\frac{m_{pc}^1(\cdot)(1-\alpha_{pc}^1) + m_{pc}^0(\cdot)(1-\alpha_{pc}^0)}{(1-\alpha_p^1)} \le \frac{m_{pc}^1(\cdot)\alpha_{pc}^1 + m_{pc}^0(\cdot)\alpha_{pc}^0}{\alpha_p^1}$$

(after m_p^1 and denominator τ terms cancel out), which rearranges to

$$(1-\tau)\phi\kappa_0(\phi)(\alpha_p^1 - \alpha_{pc}^0) \le \tau\kappa_1(s^*(\phi), \beta)\alpha_c(1-\alpha_p^1) \tag{9}$$

We want to identify an upper bound on the LHS, and a lower bound on the RHS, each holding uniformly over $\phi \in [0, \hat{\phi}]$, so as to (implicitly) define the $\bar{\alpha}_p^1$ threshold (from Assumption 1 (v)) below which (9) is guaranteed to hold. First observe that if $\alpha_p^1 \le \alpha_{pc}^0$, then the LHS of (9) is nonpositive, and the RHS is nonnegative, so (9) is satisfied. Thus the remainder of our derivation of $\bar{\alpha}_p^1$ will restrict attention to the case of $\alpha_p^1 > \alpha_{pc}^0$.

For the LHS upper bound: $\phi\kappa_0(\phi)$ is maximized at $\hat{\phi}$, since each function in the composition $F \circ k_c^* \circ q^*(\phi)$ is weakly increasing. So LHS $\le (1-\tau)\hat{\phi}\kappa_0(\hat{\phi})(\alpha_p^1 - \alpha_{pc}^0)$.

For the RHS lower bound:

$$\text{Let} \quad \underline{\ddot{k}}_c^1 := \alpha_c(1-\alpha_p^1) + (1-\lambda)\alpha_c(1-\alpha_p^1)(s_0 - \beta) - \beta\lambda, \quad \text{and let} \quad \underline{\kappa}_1 := F(\underline{\ddot{k}}_c^1)$$

We will show that $\underline{\kappa}_1 \le \kappa_1(s^*(\phi), \beta)$ for all $\phi \in [0, \hat{\phi}]$, and that $\underline{\kappa}_1 > 0$.

- Observe that $\ddot{k}_c^1(s, \beta) = \alpha_c(1-\alpha_p^1) + J_1 s + J_0\beta = \alpha_c(1-\alpha_p^1) + J_1(s-\beta) - \lambda\beta$, where:
  - $J_1 = \alpha_{pc}^1(1-\lambda) - \alpha_p^1 = (1-\lambda)\alpha_c(1-\alpha_p^1) - \lambda\alpha_p^1 < (1-\lambda)\alpha_c(1-\alpha_p^1)$

- $J_0 = (1-\alpha_{pc}^1)(1-\lambda) - (1-\alpha_p^1) = -\lambda - J_1$

- If $J_1 \geq 0$, then $J_1(s-\beta)$ is increasing in s, and thus for all $s \in \left[s^*(0), s^*(\hat{\phi})\right] = [s_0, \beta]$, we have $\ddot{k}_c^1(s,\beta) \geq \ddot{k}_c^1(s_0,\beta) = \alpha_c(1-\alpha_p^1) + J_1(s_0-\beta) - \beta\lambda > \underline{\ddot{k}}_c^1$.

- If $J_1 < 0$, then $\ddot{k}_c^1(s,\beta) \geq \alpha_c(1-\alpha_p^1) - \beta\lambda > \underline{\ddot{k}}_c^1$.

In both cases, $\underline{\ddot{k}}_c^1$ provides a lower bound for $\ddot{k}_c^1(s^*(\phi),\beta)$ which holds uniformly over $\phi \in [0,\hat{\phi}]$. To see that $\underline{\kappa}_1 > 0$, we invoke the fact that, per Assumption 1 (vi),

$$\min_{\sigma_A \in [0,\beta]^4} \tilde{k}_c^1 = \tilde{k}_c^1(s=t=\beta, q=v=0) = k_p + \alpha_c - \alpha_p^1 - \beta > \underline{k_c}$$

and we will show that $\underline{\ddot{k}}_c^1 > k_p + \alpha_c - \alpha_p^1 - \beta$:

$$\begin{aligned} \underline{\ddot{k}}_c^1 - \left(k_p + \alpha_c - \alpha_p^1 - \beta\right) &= \alpha_c(1-\alpha_p^1) - \alpha_c + \alpha_p^1 - k_p + (1-\lambda)\alpha_c(1-\alpha_p^1)(s_0-\beta) - \beta\lambda + \beta \\ &= \left[\alpha_p^1(1-\alpha_c) - k_p\right] + (1-\lambda)\left[\beta + \alpha_c(1-\alpha_p^1)(s_0-\beta)\right] > 0 \end{aligned}$$

The quantity in the first bracket is positive by Assumption 1 (iii), and the quantity in the second bracket is positive by the fact that $\alpha_c(1-\alpha_p^1)(\beta - s_0) < \beta - s_0 < \beta$.

Plugging in the two bounds derived thus far: a sufficient condition for (9) to hold for all $\phi \in [0,\hat{\phi}]$ is

$$(1-\tau)\hat{\phi}\kappa_0(\hat{\phi})(\alpha_p^1 - \alpha_{pc}^0) \leq \tau\underline{\kappa}_1\alpha_c(1-\alpha_p^1) \tag{10}$$

Next, we will verify that (10) is equivalent to a threshold condition $\alpha_p^1 \leq \bar{\alpha}_p^1$, with $\bar{\alpha}_p^1 \in \left(\alpha_{pc}^0, 1\right)$. Observe that $\hat{\phi}$ and $\kappa_0(\hat{\phi})$ are both constant in α_p^1. (Recall that $\hat{\phi}$ is the solution to $s^*(\phi) = \beta$; both $s^*(\phi)$ and κ_0 depend on $q^*(\phi)$, which does not depend on α_p^1.) Thus the LHS of (10) is continuous and strictly increasing in α_p^1. Further, $\underline{\kappa}_1$ is weakly decreasing in α_p^1, since $d\underline{\ddot{k}}_c^1/d\alpha_p^1 = -\alpha_c\left[1-(1-\lambda)(\beta-s_0)\right] < 0$. So the RHS of (10) is continuous and decreasing in α_p^1. Thus there exists at most one value of α_p^1, denoted $\bar{\alpha}_p^1$, that satisfies (10) at equality. Finally, evaluated at $\alpha_p^1 = \alpha_{pc}^0$, we have LHS $= 0 <$ RHS;[23] and at $\alpha_p^1 = 1$, LHS $> 0 =$ RHS; so $\bar{\alpha}_p^1$ exists, and satisfies $\bar{\alpha}_p^1 \in \left(\alpha_{pc}^0, 1\right)$.

Solving for ϕ^*: Our equilibrium of interest divides into two cases: (i) $s^*(\phi) = \beta$, $q^*(\phi) = q^\sharp$; or (ii) $s^*(\phi) \in (0,\beta), q^*(\phi) \in [0, q^\sharp)$. Which case can be supported depends on the sign of $M(\phi)$:

$$\begin{aligned} \text{Let} \quad M(\phi) &:= \frac{\pi}{1-\pi}\tau\alpha_p^1 - \tau\left[\kappa_1(s^*(\phi),\beta)\alpha_{pc}^1(1-\lambda) + (1-\kappa_1(s^*(\phi),\beta))\alpha_p^1\right] - (1-\tau)\kappa_0(\phi)\phi(1-\lambda)\alpha_{pc}^0 \\ &= \frac{\pi}{1-\pi}\tau\alpha_p^1 - \tau\left[\kappa_1(s^*(\phi),\beta)J_1 + \alpha_p^1\right] - (1-\tau)\kappa_0(\phi)\phi(1-\lambda)\alpha_{pc}^0 \end{aligned} \tag{11}$$

with $\kappa_1(s,t)$, $\kappa_0(\phi)$, and J_1 defined above. Observe that $M(\phi) > 0 \iff \mu^{110}(\phi) > \frac{1}{2}$. Further:

- $\kappa_0(\phi)$ is continuous and weakly increasing in ϕ (via composition of $F \circ k_c^* \circ q^*(\phi)$, each step of which is continuous and weakly increasing).

- κ_1 is continuous and monotonic in ϕ, with the sign of the derivative equal to the sign of J_1 (again by the composition $F \circ \ddot{k}_c^1 \circ s^*(\phi)$, with $\dfrac{\partial \kappa_1}{\partial s} = F'(\ddot{k}_c^1(s,\beta))[\alpha_{pc}^1(1-\lambda) - \alpha_p^1]$). Since $\kappa_1(s^*(\phi),\beta)$

[23] We established above that $\underline{\kappa}_1 > 0$ evaluated at the actual value of α_p^1, which is restricted to $\alpha_p^1 > \alpha_{pc}^0$ at this point in the analysis; since $\underline{\ddot{k}}_c^1$ is strictly decreasing in α_p^1, it follows that $\underline{\ddot{k}}_c^1|_{\alpha_p^1=\alpha_{pc}^0} > \underline{\ddot{k}}_c^1$, and hence $\underline{\kappa}_1|_{\alpha_p^1=\alpha_{pc}^0} > 0$.

only enters $M(\phi)$ multiplied by a J_1 coefficient, we have that the expression $\kappa_1(s^*(\phi), \beta)J_1$ is increasing in ϕ, with derivative $F'(\ddot{k}_c^1(s,\beta))J_1^2 \dfrac{ds^*}{d\phi} \geq 0$.

- $\kappa_0(\phi) > 0$, following Remark 7.

Thus $M(\phi)$ is continuous and strictly decreasing in ϕ (strictness due to the $\kappa_0(\phi) > 0$ term multiplied by ϕ). In addition, we know that $M(0) > 0$ by A1(iv), as shown in the proof of Lemma 8.[24]

We evaluate the two cases of $M(\hat{\phi}) \geq 0$ and $M(\hat{\phi}) < 0$, and derive the equilibrium value ϕ^* from each. We will see that in the $M(\hat{\phi}) \geq 0$ case, an equilibrium exists with $\phi^* = \hat{\phi}$; and in the $M(\hat{\phi}) < 0$ case, an equilibrium exists with ϕ^* as the interior root solving $M(\phi) = 0$.

First, suppose that $M(\hat{\phi}) \geq 0$. Then $\mu^{110}(\hat{\phi}) \geq \frac{1}{2}$, and the audience's strategy of $s = \beta$ is a best-response given their beliefs which derive from the conjectured leader strategy $\hat{\phi}$. In addition, $s = \beta$ satisfies $s = s^*(\hat{\phi})$ (by definition of $\hat{\phi}$), which means $\widehat{k}_p^0(\sigma_A^*(\hat{\phi})) = k_p$, so the leader is indifferent in ϕ, making $\phi = \hat{\phi}$ incentive-compatible. And because of (10), we know that $\mu^{110} \geq \frac{1}{2} \implies \mu^{100} \geq \frac{1}{2}$, which supports $t = \beta$. Thus if $M(\hat{\phi}) \geq 0$ (and the boundary condition in $\tilde{k}_p^1$ is satisfied, as shown below), a CSRE is supported in which $\phi = \phi^* = \hat{\phi}$, $q = q^*(\phi^*) = q^\sharp$, $s = s^*(\phi^*) = \beta$, and $t = v = \beta$.

Alternatively, suppose $M(\hat{\phi}) < 0$. Since $M(0) > 0$ and M is continuous in ϕ, by the intermediate value theorem, there exists a $\phi^* \in (0, \hat{\phi})$ that solves $M(\phi^*) = 0$. By strict monotonicity of M, we know that this ϕ^* is unique. To fully characterize the equilibrium, we have $\mu^{110} = \frac{1}{2}$, which again implies $\mu^{100} \geq \frac{1}{2}$ by (10), supporting $t = \beta$; and $s = s^*(\phi^*) \in [s_0, \beta)$, and $q = q^*(\phi^*) \in [0, q^\sharp)$.

Verifying $\tilde{k}_p^1(\sigma_A^*(\phi^*)) > k_p$: It remains to show that in our candidate equilibrium (featuring ϕ^* and $\sigma_A^*(\phi^*)$), playing $a_p = 1$ when $\omega = 1$ and $k_c > \ddot{k}_c^1$ is incentive-compatible for the leader. This is satisfied by $\tilde{k}_p^1(\sigma_A^*(\phi^*)) > k_p$. Plugging in the equilibrium values of σ_A^*, this condition becomes:

$$
\begin{aligned}
k_p <& \tilde{k}_p^1(t = v = \beta, q^*, s^*) = \alpha_p^1 - \alpha_0 + \alpha_p^1 s^* + (1-\alpha_p^1)\beta - \alpha_0 q^* - (1-\alpha_0)\beta \\
&= (\alpha_p^1 - \alpha_0)(1-\beta) + \frac{\alpha_p^1}{\alpha_{pc}^0}\left[\frac{k_p - \alpha_p^0(1-\alpha_c)}{1-\lambda} + \alpha_p^0(1-\alpha_c)\beta + \alpha_c q^*\right] - \alpha_0 q^* \qquad (12)
\end{aligned}
$$

We can see that the RHS is increasing in λ, and that $\left(\frac{\alpha_c \alpha_p^1}{\alpha_{pc}^0} - \alpha_0\right) q^* > 0$ (see Remark 8); so plugging in $\lambda = 0, q^* = 0$ gives a lower bound on the RHS. Thus $\tilde{k}_p^1(\sigma_A^*(\phi^*)) > k_p$ is satisfied if:

$$
\begin{aligned}
k_p &< (\alpha_p^1 - \alpha_0)(1-\beta) + \frac{\alpha_p^1}{\alpha_{pc}^0}\left[k_p - \alpha_p^0(1-\alpha_c) + \alpha_p^0(1-\alpha_c)\beta\right] \\
k_p(\alpha_{pc}^0 - \alpha_p^1) &< \alpha_{pc}^0(\alpha_p^1 - \alpha_0)(1-\beta) - \alpha_p^1\alpha_p^0(1-\alpha_c)(1-\beta) = (1-\beta)\left[\alpha_p^1\alpha_c - \alpha_{pc}^0\alpha_0\right] \qquad (13)
\end{aligned}
$$

This is satisfied given the upper bound on β in Assumption 1 (i) (where the term in square brackets in the RHS of (13) is positive, per Remark 8). ■

Comment: The restriction that $\alpha_p^1 \leq \bar{\alpha}_p^1$ simplifies our search for interior CSRE, by ensuring that $t = \beta$ is a best response given $\phi^*, s^*(\phi^*), q^*(\phi^*)$. If $\alpha_p^1 \leq \bar{\alpha}_p^1$ were violated, it is possible that at the

[24] To see why the Lemma 8 result carries over: Observe that $M(\phi)$ is decreasing in $s^*(\phi)$, again because $J_1 \frac{\partial \kappa_1}{\partial s} = F'(\cdot)J_1^2 \geq 0$. Let $\tilde{M}(\phi)$ denote $M(\phi)$ with $s = \beta$ replacing $s = s^*(\phi)$ in the κ_1 terms; so $\tilde{M}(\phi) \leq M(\phi)$ for fixed ϕ. The proof of Lemma 8 showed that $\tilde{M}(\phi = 0) > 0$ (by A1(iv)), which gave us $\mu^{110} > \frac{1}{2}$ in the NCRE.

$\phi^*, s^*(\phi^*)$ values derived in the proof of Lemma 9, we would have $\mu^{100}(s^*(\phi^*), t = \beta, \phi^*) < \frac{1}{2}$, making $t = \beta$ not a best response; in this case we would need to search for interior CSRE featuring $t < \beta$.

For substantive intuition as to why $\alpha_p^1 \leq \bar{\alpha}_p^1$ is sufficient to support the interior CSRE characterized in Lemma 9: if public action alone ($a = (1, 0)$) is very likely to succeed when $\omega = 1$ (that is, if α_p^1 is high), then the audience's observation of public action with failure leads them to infer that the public action was very likely part of a cover story ($a = (1, 1)$ when $\omega = 0$).

Remark 10 *All interior RE satisfy* $k_p < \widehat{k}_p^1$.

Proof of Remark 10: Interior RE are, by definition, partitioned into NCRE and CSRE. If an interior RE is CSRE, then Lemma 4 gives $k_p \leq \widehat{k}_p^0$ and Remark 5 gives $\widehat{k}_p^0 < \widehat{k}_p^1$ (for any fixed σ_A). If an interior RE is NCRE, then Lemma 8 gives $k_p < \widehat{k}_p^1$. (In the NCRE case of $k_p = \widehat{k}_p^{NCRE}$, recall that $\widehat{k}_p^{NCRE} = \widehat{k}_p^0$ evaluated at the equilibrium $\sigma_A = (s = t = v = \beta, q = q^*(0))$, which again is $< \widehat{k}_p^1$.) ■

Proposition 2 (RE existence) *An interior RE always exists.*

Proof of Proposition 2: Follows directly from Lemmas 8 and 9. Recall that $\widehat{k}_p^{NCRE}$ is a function of primitives (via $q^*(0)$, which is uniquely pinned by primitives). From Lemma 8, an interior NCRE exists if $\widehat{k}_p^{NCRE} \leq k_p$. From Lemma 9, an interior CSRE exists if $\widehat{k}_p^{NCRE} > k_p$. ■

Lemma 10 (Monotonicity of $\boldsymbol{\widehat{k}_p^{NCRE}(\lambda)}$) *Let* $\widehat{k}_p^{NCRE}(\lambda) = \alpha_p^0(1 - \alpha_c) + (1 - \lambda)\alpha_c[\beta - q^*(0; \lambda)]$. *Then* $\widehat{k}_p^{NCRE}(\lambda)$ *is continuous and weakly decreasing in* λ*, and strictly decreasing for any* λ *such that* $q^*(0; \lambda) < \beta$.

Proof of Lemma 10: Consider the three branches of $q^*(0; \lambda)$ given in Lemma 7:

- If $q^*(0; \lambda) = 0$, then λ enters linearly in the expression of $\widehat{k}_p^{NCRE}$, with coefficient $-\alpha_c \beta$.
- $\widehat{k}_p^{NCRE}$ is flat in any interval of λ for which $q^*(0; \lambda) = \beta$.
- If $q^*(0; \lambda) \in (0, \beta)$, then it suffices to show that $q^*(0; \lambda)$ is continuous and non-decreasing in λ. We will do so via implicit differentiation at the equilibrium condition $\mu^{010}(q, \phi = 0; \lambda) = \frac{1}{2}$, per Lemma 7. Let $\Gamma(\lambda) = \alpha_c(1 - \lambda) - \alpha_0$, and observe $\Gamma' = -\alpha_c < 0$ (with $\Gamma > 0$ given by A1(ii)). Also observe that $k_c^*(q, v = \beta; \lambda) = \alpha_c - \alpha_0 + q\Gamma(\lambda) + \beta(-\lambda - \Gamma(\lambda))$ is strictly decreasing in λ and strictly increasing in q. λ and q only enter μ^{010} via the $F(k_c^*)\Gamma(\lambda)$ term in the denominator; $F(k_c^*)\Gamma(\lambda)$ is strictly decreasing in λ, strictly increasing in q (as F has positive density at k_c^*, per A1(vi)), and continuous in both arguments. So:

$$\frac{\partial q^*(0; \lambda)}{\partial \lambda} = -\frac{\partial \mu^{010}(q, 0; \lambda)/\partial \lambda}{\partial \mu^{010}(q, 0; \lambda)/\partial q} = -\frac{(+)}{(-)} > 0$$

Altogether, we have that $\widehat{k}_p^{NCRE}(\lambda)$ is weakly decreasing in λ, and strictly decreasing on any interval over which $q^*(0; \lambda) < \beta$; and we have continuity within each of the three cases. Continuity across the three cases holds because at each boundary of the interior branch, $q^*(0; \lambda)$ agrees with the respective corner branch (as shown in Lemma 7). ■

Proposition 3 (Transparency threshold for cover stories among interior RE) *If $\widehat{k}_p^{NCRE}(0) > k_p$, define $\bar{\lambda}$ as the unique implicit solution to $\widehat{k}_p^{NCRE}(\bar{\lambda}) = k_p$, or equivalently, to $\bar{\lambda} = 1 - \left(\dfrac{k_p - \alpha_p^0(1-\alpha_c)}{\alpha_c(\beta - q^*(0;\bar{\lambda}))}\right)$. Otherwise, if $\widehat{k}_p^{NCRE}(0) \leq k_p$, let $\bar{\lambda} = 0$. Then:*

- *If $\lambda > \bar{\lambda}$, all interior RE are NCRE.*[25]
- *If $\lambda < \bar{\lambda}$, all interior RE are CSRE.*

Proof of Proposition 3: First observe the following, for the case of $\widehat{k}_p^{NCRE}(0) > k_p$:

- $\widehat{k}_p^{NCRE}(\lambda) = \alpha_p^0(1-\alpha_c) < k_p$ for any λ for which $q^*(0;\lambda) = \beta$. So any solution to $\widehat{k}_p^{NCRE}(\lambda) = k_p$ must satisfy $q^*(0;\lambda) < \beta$.
- For sufficiently large λ, we have $q^*(0;\lambda) = \beta$ (by Lemma 7), and thus $\widehat{k}_p^{NCRE}(\lambda) < k_p$. Since $\widehat{k}_p^{NCRE}(0) > k_p$, continuity of $\widehat{k}_p^{NCRE}(\lambda)$ implies that a solution to $\widehat{k}_p^{NCRE}(\lambda) = k_p$ exists.
- Then uniqueness of the solution $\bar{\lambda}$ follows from strict monotonicity of $\widehat{k}_p^{NCRE}(\lambda)$, per Lemma 10.

Thus we have a unique solution $\bar{\lambda}$, with $\text{sign}(\lambda - \bar{\lambda}) = \text{sign}(k_p - \widehat{k}_p^{NCRE}(\lambda))$. Lemma 8 then implies that the interior NCRE is unique (among interior RE) when $\lambda > \bar{\lambda}$; that an interior NCRE exists when $\lambda \geq \bar{\lambda}$; and that no interior NCRE exists when $\lambda < \bar{\lambda}$.

Proposition 2 showed that an interior RE always exists. The set of interior RE is, by definition, partitioned into NCRE and CSRE. Since no interior NCRE exists when $\lambda < \bar{\lambda}$, that means an interior CSRE exists when $\lambda < \bar{\lambda}$ (or equivalently, when $k_p < \widehat{k}_p^{NCRE}(\lambda)$, per Lemma 9).

Finally, if $\widehat{k}_p^{NCRE}(0) \leq k_p$, the monotonicity result from Lemma 10 implies $\widehat{k}_p^{NCRE}(\lambda) < k_p$ for all $\lambda > 0$, and therefore that all interior RE are NCRE. (The strict inequality $\widehat{k}_p^{NCRE}(\lambda) < k_p$ follows from the fact that for any $\lambda > 0$, either (i) $q^*(0;\lambda) = \beta$, so $\widehat{k}_p^{NCRE}(\lambda) < k_p$; or (ii) $q^*(0;\lambda) < \beta$, so monotonicity of $\widehat{k}_p^{NCRE}(\lambda)$ is strict.) ■

Proposition 4 (Transparency threshold for non-punishment among interior RE) *There exists a threshold $\bar{\bar{\lambda}} \in \left[\bar{\lambda}, \frac{\alpha_c - \alpha_0}{\alpha_c}\right)$—with $\bar{\bar{\lambda}} = \bar{\lambda}$ only if $\bar{\bar{\lambda}} = \bar{\lambda} = 0$—such that, in any interior RE, the audience punishes unexplained success (that is, plays $q < \beta$) iff $\lambda < \bar{\bar{\lambda}}$.*

Proof of Proposition 4: We want to find $\bar{\bar{\lambda}}$, the lowest λ value for which $q^*(\phi^*;\lambda) = \beta$, at the equilibrium ϕ^* of an interior RE. (Remark 10 showed that all interior RE feature $k_p < \widehat{k}_p^1$, so the results of Lemma 7 apply—specifically, that $v = \beta$, and $q^*(\phi;\lambda)$ is the uniquely-defined equilibrium value of q.)

We first want to establish that in any interior RE, $q^*(\phi^*;\lambda) = \beta \implies \lambda > \bar{\lambda}$. If $\bar{\lambda} = 0$, then $\lambda > \bar{\lambda}$ is satisfied automatically (as $\lambda \leq 0$ is inadmissible). Otherwise, if $\bar{\lambda} > 0$: suppose (toward a contradiction) that for some $\lambda \leq \bar{\lambda}$, we have $q^*(\phi^*;\lambda) = \beta$. Then $\widehat{k}_p^0(v = \beta, q = q^*(\phi^*;\lambda) = \beta) < k_p$ for any s, t (Remark 6), which means that the equilibrium must be an NCRE (by Lemma 4), with $\phi^* = 0$ and $q^*(\phi^*;\lambda) = q^*(0;\lambda) = \beta$; and thus $\widehat{k}_p^{NCRE}(\lambda) = \widehat{k}_p^0(q = v = s = t = \beta;\lambda) < k_p$. But recall (from the proof of Proposition 3) that $\text{sign}(\lambda - \bar{\lambda}) = \text{sign}(k_p - \widehat{k}_p^{NCRE}(\lambda))$; so $\widehat{k}_p^{NCRE}(\lambda) < k_p$ contradicts $\lambda \leq \bar{\lambda}$. Therefore, $q^*(\phi^*;\lambda) = \beta \implies \lambda > \bar{\lambda}$, and $\lambda > \bar{\lambda} \implies \phi^* = 0$.

From Lemma 7, we know that $q^*(0;\lambda) = \beta \iff \mu^{010}(q = \beta, \phi = 0; \lambda) \geq \frac{1}{2}$. So we want to find the lowest λ value for which $\mu^{010}(\beta, 0; \lambda) \geq \frac{1}{2}$ is satisfied. Observe the following:

[25] If $\lambda = \bar{\lambda}$, an interior NCRE exists, but may co-exist with a CSRE.

- $\mu^{010}(q=\beta, \phi=0; \lambda)$ is continuous and strictly increasing in λ: λ only enters $\mu^{010}(\beta, 0; \lambda)$ via the $F(k_c^*(\beta,\beta))[\alpha_c(1-\lambda)-\alpha_0]$ term in the denominator, where $k_c^*(\beta,\beta)=\alpha_c-\alpha_0-\beta\lambda$, and both $F(k_c^*)>0$ (by Remark 7) and $[\alpha_c(1-\lambda)-\alpha_0]>0$ (by A1(ii)).

- As $\lambda \to \frac{\alpha_c-\alpha_0}{\alpha_c}$ (the limiting value allowed by A1(ii)), we have $\mu^{010}(\beta, 0; \lambda) \to \pi > \frac{1}{2}$.

Then there are two cases to consider: $\lim_{\lambda\to 0^+} \mu^{010}(\beta, 0; \lambda) \geq \frac{1}{2}$; and $\lim_{\lambda\to 0^+} \mu^{010}(\beta, 0; \lambda) < \frac{1}{2}$.

- If $\lim_{\lambda\to 0^+} \mu^{010}(\beta, 0; \lambda) \geq \frac{1}{2}$, this means that $q=\beta$ for all admissible λ (by monotonicity of $\mu^{010}(\beta, 0; \lambda)$), and thus $\bar{\bar{\lambda}}=0$. It also means that $\bar{\lambda}=0$ (since, as established above, $q^*(\phi^*;\lambda)=\beta \implies \lambda>\bar{\lambda}$). So in this case, $\lambda > \bar{\bar{\lambda}} \implies q=\beta$ (and the other half of the "iff" statement is vacuous, as $\lambda \leq \bar{\bar{\lambda}} = 0$ is not admissible).

- If $\lim_{\lambda\to 0^+} \mu^{010}(\beta, 0; \lambda) < \frac{1}{2}$, then either $\bar{\lambda}=0$; or $\bar{\lambda}>0$, so by Proposition 3, we have $q^*(0;\bar{\lambda})<\beta$, and thus $\mu^{010}(\beta,0;\bar{\lambda}) < \frac{1}{2}$ (by Lemma 7). In either case, $\lim_{\lambda\to\bar{\lambda}^+} \mu^{010}(\beta,0;\lambda) < \frac{1}{2}$. Given the continuity and strict monotonicity of $\mu^{010}(\beta,0;\lambda)$, and the upper endpoint noted above, we can thus apply the intermediate value theorem: there exists a unique $\bar{\bar{\lambda}} \in \left(\bar{\lambda}, \frac{\alpha_c-\alpha_0}{\alpha_c}\right)$ satisfying $\mu^{010}(\beta,0;\bar{\bar{\lambda}})=\frac{1}{2}$. By strict monotonicity of $\mu^{010}(\beta,0;\lambda)$ in λ, $\text{sign}(\mu^{010}(\beta,0;\lambda)-\frac{1}{2}) = \text{sign}(\lambda-\bar{\bar{\lambda}})$; thus $\lambda \geq \bar{\bar{\lambda}} \iff q^*(\phi^*;\lambda)=\beta$.

■

Proposition 1, restated (Interior RE regions, partitioned by $\boldsymbol{\lambda}$) *Within the class of interior RE, there exist thresholds $\bar{\lambda}$ and $\bar{\bar{\lambda}}$, with $\bar{\lambda} \leq \bar{\bar{\lambda}}$ (and $\bar{\lambda}=\bar{\bar{\lambda}}$ only if $\bar{\lambda}=\bar{\bar{\lambda}}=0$), such that:*

- *If transparency is high ($\lambda \geq \bar{\bar{\lambda}}$), the audience does not punish the leader in the absence of direct exposure of covert action.*

- *If transparency is in an intermediate range ($\bar{\lambda} < \lambda < \bar{\bar{\lambda}}$), the leader does not use a cover story, and the audience may punish the leader for an "unexplained success".*

- *If transparency is low ($\lambda < \bar{\lambda}$), the leader uses a cover story with positive probability.*

Proof of Proposition 1: Follows directly from Propositions 3 and 4. ■

Corollary 1, restated (Comparative statics on CSRE transparency threshold) *If $q^*(0;\bar{\lambda}) < \beta$ (the non-trivial case from Proposition 3), then the $\bar{\lambda}$ threshold is:*

- *increasing in α_c;*

- *decreasing in k_p;*

- *increasing in β when $q^*(0;\bar{\lambda}(\beta))=0$, and decreasing in β when $q^*(0;\bar{\lambda}(\beta)) \in (0,\beta)$. Therefore:*
  - *If the set $\{\beta : q^*(0;\bar{\lambda}(\beta))=0\}$ is nonempty, let $\hat{\beta} := \sup\{\beta : q^*(0;\bar{\lambda}(\beta))=0\}$. If $\hat{\beta}$ is interior to the admissible range of β, then $\bar{\lambda}$ is single-peaked in β—increasing for $\beta<\hat{\beta}$ and decreasing for $\beta>\hat{\beta}$; otherwise, $\bar{\lambda}$ is increasing in β throughout the admissible range.*
  - *If α_0 is sufficiently low, $\bar{\lambda}$ is increasing in β for any admissible β.*

Proof of Corollary 1: Define $H(\lambda, x) = \widehat{k}_p^{NCRE}(\lambda, x) - k_p = \alpha_p^0(1-\alpha_c) + (1-\lambda)\alpha_c(\beta - q^*(0;\lambda)) - k_p$ for parameters $x = k_p, \alpha_c, \beta$. Then $\bar{\lambda}$ is the solution to $H(\bar{\lambda}, x) = 0$, or $\bar{\lambda} = 1 - \left(\dfrac{k_p - \alpha_p^0(1-\alpha_c)}{\alpha_c(\beta - q^*(0;\bar{\lambda}))}\right)$.

There are two cases to consider for the comparative statics analysis: $q^*(0;\bar{\lambda}) = 0$, and $q^*(0;\bar{\lambda}) \in (0,\beta)$. If $q^*(0;\bar{\lambda}) = 0$, then the comparative statics are simply partial derivatives:

$$\frac{\partial\bar{\lambda}}{\partial\alpha_c} = -\partial\left(\frac{k_p - \alpha_p^0}{\alpha_c\beta}\right)/\partial\alpha_c > 0, \quad \frac{\partial\bar{\lambda}}{\partial k_p} < 0, \quad \frac{\partial\bar{\lambda}}{\partial\beta} > 0$$

If $q^*(0;\bar{\lambda}) \in (0,\beta)$, then we can compute $\frac{d\bar{\lambda}}{dx} = \frac{-\partial H/\partial x}{\partial H/\partial\lambda}\Big|_{\lambda=\bar{\lambda}}$ for $x = k_p, \alpha_c, \beta$. Let $\Gamma(\lambda) = \alpha_c(1-\lambda) - \alpha_0$. Note that λ and q^* only enter $\mu^{010} = \frac{\pi\alpha_0}{\pi\alpha_0 + (1-\pi)[\alpha_0 + F(k_c^*)\Gamma(\lambda)]}$ through the $F(k_c^*)\Gamma(\lambda)$ term in the denominator; and $F(k_c^*)\Gamma(\lambda)$ is increasing in q^*, and decreasing in λ. So then we have:

$$\frac{\partial H}{\partial\lambda} = -\alpha_c(\beta - q^*) - (1-\lambda)\alpha_c\frac{\partial q^*}{\partial\lambda}, \quad \text{where} \quad \frac{\partial q^*}{\partial\lambda} = \frac{-\partial\mu^{010}/\partial\lambda}{\partial\mu^{010}/\partial q^*} = -\frac{(+)}{(-)} > 0, \quad \text{so} \quad \frac{\partial H}{\partial\lambda} < 0$$

Considering each parameter $x = k_p, \alpha_c, \beta$ in turn:

- k_p: $\frac{\partial H}{\partial k_p} = -1$, so $\frac{d\bar{\lambda}}{dk_p} = -\frac{(-)}{(-)} < 0$.
- α_c: $\frac{\partial H}{\partial\alpha_c} = -\alpha_p^0 + (1-\bar{\lambda})(\beta - q^*) - (1-\bar{\lambda})\alpha_c\frac{\partial q^*}{\partial\alpha_c}$
  - $\frac{\partial q^*}{\partial\alpha_c} = -\frac{\partial\mu^{010}/\partial\alpha_c}{\partial\mu^{010}/\partial q^*} = -\frac{(-)}{(-)} < 0$: α_c enters μ^{010} only through $F(k_c^*)\Gamma$ in the denominator, with $F(k_c^*)\Gamma$ increasing in α_c.
  - $H = 0$ rearranges to $-\alpha_p^0 + (1-\bar{\lambda})(\beta - q^*) = \frac{k_p - \alpha_p^0}{\alpha_c}$, which is > 0; substituting, we get

$$\frac{\partial H}{\partial\alpha_c} = \frac{k_p - \alpha_p^0}{\alpha_c} - (1-\bar{\lambda})\alpha_c\frac{\partial q^*}{\partial\alpha_c} > 0, \quad \text{so} \quad \frac{d\bar{\lambda}}{d\alpha_c} = -\frac{\partial H/\partial\alpha_c}{\partial H/\partial\lambda} = -\frac{(+)}{(-)} > 0$$

- β: $\frac{\partial H}{\partial\beta} = (1-\bar{\lambda})\alpha_c\left(1 - \frac{\partial q^*}{\partial\beta}\right)$
  - β enters μ^{010} only through $F(k_c^*)$, so the equilibrium condition $\mu^{010} = \frac{1}{2}$ can be rearranged to an expression $k_c^*(q^*(\phi), v = \beta) = F^{-1}(\xi)$, where ξ is constant in β. Then:

$$\frac{\partial q^*}{\partial\beta} = -\frac{\partial k_c^*/\partial\beta}{\partial k_c^*/\partial q^*} = -\frac{(-\lambda - \Gamma)}{\Gamma} = \frac{\lambda + \Gamma}{\Gamma} > 1, \quad \text{so} \quad \frac{\partial H}{\partial\beta} < 0, \quad \frac{d\bar{\lambda}}{d\beta} < 0$$

Altogether, across both cases of $q^*(0;\bar{\lambda}) = 0$ and $q^*(0;\bar{\lambda}) \in (0,\beta)$: $\bar{\lambda}$ is decreasing in k_p and increasing in α_c; and increasing in β when $q^*(0;\bar{\lambda}) = 0$, and decreasing in β when $q^*(0;\bar{\lambda}) \in (0,\beta)$.

Single-peakedness of $\bar{\lambda}$ in β: Next, we must verify that these two differently-signed derivatives of $\bar{\lambda}$ with respect to β imply an interior peak rather than an interior trough. On the corner branch, $q^*(0;\bar{\lambda}) = 0$, so $\bar{\lambda}$ is increasing in β by the partial derivative above. On the interior branch, the implicit differentiation above gives $\frac{d\bar{\lambda}}{d\beta} < 0$. It remains to show that the $q^*(0;\bar{\lambda}) = 0$ branch occurs before the interior branch, if both occur. From Lemma 7, we have that $q^*(0;\bar{\lambda}) = 0 \iff \mu^{010}(q = 0, \phi = 0; \bar{\lambda}) \leq \frac{1}{2}$. We will show that $\mu^{010}(0,0;\bar{\lambda}) \leq \frac{1}{2}$ holds for β below some threshold $\hat{\beta}$.

Let $N = k_p - \alpha_p^0(1-\alpha_c)$; note $N > 0$ by A1(iii). $H = 0$ rearranges to $1 - \bar{\lambda} = \frac{N}{\alpha_c\beta}$. On the corner branch with $q^*(0;\bar{\lambda}) = 0$, we can see that $\mu^{010} = \frac{\pi\alpha_0}{\pi\alpha_0 + (1-\pi)\left[\alpha_0 + F(k_c^*)\Gamma(\bar{\lambda})\right]}$, and β only enters via the

$F(k_c^*)\Gamma(\bar\lambda)$ term. Plugging in, we get

$$\Gamma(\bar\lambda) = \frac{N}{\beta} - \alpha_0, \quad \text{and} \quad k_c^*(0,\beta) = \alpha_c - \alpha_0 + \beta\left[(1-\alpha_c)(1-\bar\lambda) - (1-\alpha_0)\right] = \alpha_c - \alpha_0 + \frac{(1-\alpha_c)N}{\alpha_c} - \beta(1-\alpha_0)$$

Both terms are decreasing in β, so $\mu^{010}(0,0;\bar\lambda)$ is increasing in β. Thus the condition $\mu^{010}(0,0;\bar\lambda) \le \frac{1}{2}$ holds on an interval $\beta \le \hat\beta$. Since $\frac{d\bar\lambda}{d\beta} > 0$ on the corner branch with $\beta \le \hat\beta$, and $\frac{d\bar\lambda}{d\beta} < 0$ on the interior branch, it follows that $\bar\lambda$ is single-peaked in β.

$\bar\lambda$ increasing in β when α_0 is low: We established above that $\bar\lambda$ is increasing in β whenever $q^*(0;\bar\lambda(\beta)) = 0$, which (by Lemma 7) holds whenever $\mu^{010}(0,0;\bar\lambda) \le \frac{1}{2}$. We want to find an upper bound for $\mu^{010}(0,0;\bar\lambda)$ which is independent of β, and show that this bound approaches zero as $\alpha_0 \to 0^+$. For this comparative-static analysis, hold fixed F (including its support $\left(\underline{k_c}, \overline{k_c}\right)$) and all other primitives as α_0 varies, and restrict attention to sufficiently small values of α_0 for which A1 continues to hold.

When $q^*(0;\bar\lambda) = 0$, we have $\bar\lambda = 1 - \frac{N}{\alpha_c\beta}$, or $\alpha_c(1-\bar\lambda) = N/\beta > N$, since $\beta < 1$ per A1(i). Plugging in $\bar\lambda$ gives $k_c^*(q=0, v=\beta; \bar\lambda(\beta)) = \alpha_c - \alpha_0 + \left(\frac{1-\alpha_c}{\alpha_c}\right)N - \beta(1-\alpha_0)$. Next, note that

$$\inf_{\lambda}\min_{\sigma_A} k_c^*(\sigma_A;\lambda) = \inf_{\lambda} k_c^*(0,\beta;\lambda) = \inf_{\lambda} \alpha_c - \alpha_0 + \beta\left[(1-\alpha_c)(1-\lambda) - (1-\alpha_0)\right] = \alpha_c - \alpha_0 + \frac{\beta}{\alpha_c}(\alpha_0 - \alpha_c) =: \underline{k}_c^*$$

where $\inf_\lambda k_c^*(0,\beta;\lambda)$ is approached as $\lambda \to \frac{\alpha_c - \alpha_0}{\alpha_c}$, per A1(ii). Since A1(vi) holds for every admissible λ, we know that $\underline{k}_c^* \ge \underline{k_c}$.

Restrict attention to $\alpha_0 < N/2$. We can see that $k_c^*(0,\beta;\bar\lambda(\beta)) - \underline{k}_c^* = \left(\frac{1-\alpha_c}{\alpha_c}\right)[N - \beta\alpha_0]$, where $N - \beta\alpha_0 > N - \alpha_0 > \frac{N}{2} > 0$. Thus $k_c^*(0,\beta;\bar\lambda(\beta)) > \underline{k_c} + \left(\frac{1-\alpha_c}{\alpha_c}\right)\frac{N}{2}$, so we have

$$\kappa_0 = F(k_c^*(0,\beta;\bar\lambda(\beta))) \;\ge\; \underline{\kappa}_0 := F\left(\underline{k_c} + \left(\frac{1-\alpha_c}{\alpha_c}\right)\frac{N}{2}\right) \;>\; 0,$$

a strictly positive bound which is uniform in α_0 and β. Altogether, we have

$$\mu^{010}(0,0;\bar\lambda(\beta)) = \frac{\pi\alpha_0}{\pi\alpha_0 + (1-\pi)\left[\alpha_0 + \kappa_0(\alpha_c(1-\bar\lambda) - \alpha_0)\right]} \le \frac{\pi\alpha_0}{\pi\alpha_0 + (1-\pi)\left[\alpha_0 + \underline{\kappa}_0\,(N - N/2)\right]}$$

The RHS is independent of β. As $\alpha_0 \to 0^+$, the numerator of the RHS goes to zero while its denominator remains bounded below by $(1-\pi)\underline{\kappa}_0 N/2 > 0$. Thus for sufficiently low α_0, $\mu^{010}(0,0;\bar\lambda(\beta)) \le \frac{1}{2}$ for every admissible β. ■

Comment: For substantive intuition behind the single-peakedness of $\bar\lambda$ in β, consider the following. Starting from low β (and assuming $\hat\beta$ is interior), increasing β raises the reputational benefit of using a cover story, weighed against the fixed cost of public action k_p; so L is willing to use a cover story under a wider range of λ values. As β continues to increase past $\hat\beta$, L's reputational incentives become strong enough that she restrains her own use of covert action to the point that the audience begins to increase q, which then reduces L's *net* reputational gain from a cover story. The low-α_0 condition implies that the audience maintains $q = 0$ at $\lambda = \bar\lambda$, shutting down the second, offsetting channel.

Corollary 2, restated (Cover stories and scrutiny) *In any interior CSRE, the audience's interim beliefs of the leader's scrupulousness after observing public action (but before observing the* *outcome) are strictly less favorable than their interim beliefs after observing no public action.*

Proof of Corollary 2: The audience's interim beliefs upon observing $a_p = 1$, but before observing y or z, satisfy

$$\mu^{a_p=1} = Pr(\theta = 1|a_p = 1) = \frac{\pi\tau}{\pi\tau + (1-\pi)\left[\tau + (1-\tau)\phi\kappa_0\right]}, \qquad \kappa_0 = F(k_c^*(q^*(\phi), v = \beta)) > 0$$

$$\mu^{a_p=0} = Pr(\theta = 1|a_p = 0) = \frac{\pi(1-\tau)}{\pi(1-\tau) + (1-\pi)(1-\tau)\left[1 - \phi\kappa_0\right]}$$

If a cover story is played with positive probability ($\phi > 0$), then $\mu^{a_p=1} < \pi < \mu^{a_p=0}$. ∎

Corollary 3 (Interim beliefs of public action effectiveness) *In any interior CSRE, audience interim beliefs of public action effectiveness satisfy* $Pr(\omega = 1|a_p = 1, z = 0) \in (\tau, 1)$

Proof of Corollary 3: Leader strategies in any interior CSRE are given by Lemma 6, with $\phi > 0$ (positive probability of $a = (1,1)$ when $\omega = 0$). Audience interim beliefs of ω, after observing $a_p = 1, z = 0$ but before observing the outcome y, satisfy

$$Pr(\omega = 1|a_p = 1, z = 0) = \frac{Pr(a_p = 1, z = 0|\omega = 1)\tau}{Pr(a_p = 1, z = 0|\omega = 1)\tau + Pr(a_p = 1, z = 0|\omega = 0)(1-\tau)}$$

To see that this posterior is strictly above the prior τ but below 1, observe:

$$\begin{aligned} Pr(a_p = 1, z = 0|\omega = 1) &= \pi + (1-\pi)[(1-\kappa_1) + \kappa_1(1-\lambda)] \\ &> (1-\pi)\kappa_0\phi(1-\lambda) = Pr(a_p = 1, z = 0|\omega = 0) > 0 \end{aligned}$$

where $\kappa_1 = F(\ddot{k}_c^1) \in (0,1)$, $\kappa_0 = F(k_c^*) \in (0,1)$, and $\phi = Pr(a_p = 1|\omega = 0, a_c = 1) > 0$. ∎

Corollary 4 (Interior probability of a cover story) *In any interior CSRE, L^0 uses a cover story with strictly interior probability: $\phi = Pr(a_p = 1|\omega = 0, a_c = 1) \in (0,1)$.*

Proof of Corollary 4: An interior CSRE, by definition, requires $\phi > 0$. Suppose $\phi = 1$ within an interior CSRE. Then beliefs follow from the bottom-center cell of Table 3. So $P_0(0,y) \leq P_1(0,y)$, and $\mu^{0y0} \geq \pi$, for $y = 0,1$; and thus $v = q = \beta$, so $\widehat{k}_p^0(q = v = \beta) < k_p$ (Remark 6), contradicting the necessary condition for a cover story to be used with positive probability (per Lemma 4). ∎

7.4.3 Across-equilibrium analysis

The previous section derived our core comparative static result—that cover stories are used if and only if transparency is low—within a particular class of equilibrium (interior RE), which always exists. In this section, we further justify our evidentiary search for cover-story behavior when λ is low by demonstrating the following (focusing on the low α_0 case, for tractability): if transparency is high ($\lambda > \bar{\lambda}$), the RE with no cover story (NCRE) payoff-dominates all other equilibria for L^1; and if transparency is low ($\lambda < \bar{\lambda}$), all equilibria are either Cover Story Equilibria, or are payoff-dominated (for L^1) by a Cover Story Responsive Equilibrium (CSRE). To outline the remainder of the proof:

- Assumption 3 imposes an equilibrium refinement which narrows the set of equilibria that we make comparisons across, by ruling out equilibria in which players condition their randomizations on continuation-payoff-irrelevant variables (following Maskin & Tirole (2004))

- Lemma 12 defines the Public Action Pandering Equilibrium (PAPE) and derives conditions under which all equilibria are PAPE. Lemma 13 does the same for PIPE.
- Lemma 11 narrows down the set of possible equilibria. Remark 11 characterizes the surviving equilibria as falling into one of three categories: RE, PAPE, or PIPE.
- Proposition 5 demonstrates that when $\lambda > \bar{\lambda}$ and α_0 is low, the interior NCRE payoff-dominates all other equilibria for the scrupulous leader.
- Lemma 14 demonstrates that whenever $\lambda < \bar{\lambda}$, all PAPE are CSE.
- Proposition 6 demonstrates that when $\lambda < \bar{\lambda}$, all equilibria are either CSE or non-CSE PIPE; and when α_0 is low, an interior CSRE payoff-dominates the non-CSE PIPE for L^1.

Assumption 3 (Refinement: mixing conditional on relevant information) *Eliminate equilibria in which leader types condition their strategies on continuation-payoff-irrelevant variables. Specifically: for any ω and any region K of covert-action costs, if $BR_\theta(\omega, k_c; \sigma_A) = \{(1,0),(0,0)\}$ for both $\theta = 0,1$ and all $k_c \in K$, then there exists a scalar $\rho^\omega \in [0,1]$ such that $Pr(a = (1,0) \mid \theta, \omega, k_c) = \rho^\omega$ for all $\theta \in \{0,1\}$ and all $k_c \in K$.*

This equilibrium refinement requires that when the two leader types are both indifferent between their options that involve no covert action—that is, between $a = (0,0)$ and $a = (1,0)$—they must randomize with the same probabilities. Intuitively, the leader types differ only in their costs of covert action; if neither type is using covert action, then there is little substantive reason to expect them to behave differently. This refinement is motivated by the "Markovian" refinement used by Maskin and Tirole (2004), but adapted to the expanded action space in our setting.[26]

All results that follow invoke Assumption 3.

Lemma 11 (Narrowing the set of possible equilibria) *All equilibria satisfy at least one of the following sets of conditions: (I) $k_p \leq \tilde{k}{}_p^1$ and $\widehat{k}_p^0 \leq k_p < \widehat{k}_p^1$; or (II) $k_p \geq \tilde{k}{}_p^1$ and $\widehat{k}_p^0 < k_p \leq \widehat{k}_p^1$.*

Proof of Lemma 11: We prove the lemma by ruling out each of the following conditions in turn: (i) $k_p \geq \widehat{k}_p^1$ and $k_p \leq \tilde{k}{}_p^0$; (ii) $k_p \geq \widehat{k}_p^1$ and $\tilde{k}{}_p^0 < k_p < \tilde{k}{}_p^1$; (iii) $k_p > \widehat{k}_p^1$ and $k_p \geq \tilde{k}{}_p^1$; (iv) $k_p \leq \widehat{k}_p^0$ and $k_p > \tilde{k}{}_p^1$; (v) $k_p < \widehat{k}_p^0$ and $k_p \leq \tilde{k}{}_p^1$. (Observe that the union of (i)–(v) is the complement of (I)∪(II).)

Among the $k_p \geq \widehat{k}_p^1$ cases:

(i) $k_p \geq \widehat{k}_p^1$ and $k_p \leq \tilde{k}{}_p^0$: Any strategy profile satisfying these conditions would imply:

$$
\begin{aligned}
&P_1(1,y) = \tau g_y(\alpha_p^1) + (1-\tau)\rho_1 g_y(\alpha_p^0), \qquad g_y(\alpha) = y\alpha + (1-y)(1-\alpha) \\
&P_0(1,y) = \tau\left[(1-\kappa_1)g_y(\alpha_p^1) + \kappa_1\psi(1-\lambda)g_y(\alpha_{pc}^1)\right] + (1-\tau)\rho_0(1-\kappa_0)g_y(\alpha_p^0) \\
&\qquad \rho_\theta = Pr(a = (1,0)|\theta, \omega = 0, k_c > \tilde{k}{}_c^0(q,v)) = 1 - Pr(a = (0,0)|\theta, \omega = 0, k_c > \tilde{k}{}_c^0(q,v)) \\
&\qquad \kappa_1 = F(\tilde{k}{}_c^1(q,s,t,v)), \quad \kappa_0 \in [0,1] \\
&\qquad \psi = Pr(a = (1,1)|\theta = 0, \omega = 1, k_c < \tilde{k}{}_c^1(\cdot)) = 1 - Pr(a = (0,1)|\theta = 0, \omega = 1, k_c < \tilde{k}{}_c^1(\cdot))
\end{aligned}
$$

[26] In Maskin and Tirole (2004), leaders only have two available actions, denoted $A = \{a, b\}$; they apply the Markovian refinement in the situation where the two leader types have the same *preferences*. In our setting leaders have four available actions, $A = \{0,1\}^2$; Assumption 3 applies in the situation where the two leader types have the same *best response set* (but potentially a different preference ordering among the actions that are not in the best response set).

with $\psi = 0$ for $k_p > \widehat{k}_p^1$, and $\rho_\theta = 1$ for $k_p < \overset{\sim 0}{k_p}$. Assumption 3 requires $\rho_1 = \rho_0 = \rho$. Then we can see that $P_0(1,0) \leq P_1(1,0)$, so $\mu^{100} \geq \pi$ and $t = \beta$.

If $k_p > \widehat{k}_p^1$, then $\psi = 0$, and we can see $P_0(1,1) \leq P_1(1,1)$ for any $\kappa_1, \kappa_0 \in [0,1]$. So $s = \beta$.

Otherwise, we can see that $\frac{\pi}{1-\pi} P_1(1,1) - P_0(1,1)$ is increasing in ρ, κ_0 and decreasing in ψ. Further, given $k_p = \widehat{k}_p^1$, we have $\overset{\sim 1}{k_c} = \ddot{k}_c^1$ (Remark 5). Plug in $\rho = 0, \kappa_0 = 0, \psi = 1$, and we see that $\frac{\pi}{1-\pi} P_1(1,1) > P_0(1,1)$ is the same inequality which was shown in Lemma 8 (proving NCRE existence) to hold for any s (given $t = \beta$). Thus we have $\mu^{110} > \frac{1}{2}$, so $s = \beta$; and $\widehat{k}_p^1(s = t = \beta) > k_p$ for any q, v (Remark 6), contradicting $k_p \geq \widehat{k}_p^1$.

(ii) $k_p \geq \widehat{k}_p^1$ and $\overset{\sim 0}{k_p} < k_p < \overset{\sim 1}{k_p}$ Beliefs in this equilibrium would be constructed from the same $P_\theta(a_p, y)$ expressions as in the previous case (i), but here with $\rho = 0$ (meaning both types never play $a_p = 1$ when $\omega = 0$, since $k_p > \overset{\sim 0}{k_p}$). By the same steps, we reach the conclusion that $s = t = \beta$, with $\widehat{k}_p^1(s = t = \beta) > k_p$ contradicting $k_p \geq \widehat{k}_p^1$.

(iii) $k_p > \widehat{k}_p^1$ and $k_p \geq \overset{\sim 1}{k_p}$: Either:

- $a_p = 1$ is never reached on-path, in which case $\mu^{1y0} = \pi$ by A2, so $s = t = \beta$ and $\overset{\sim 1}{k_p}(s = t = \beta) > k_p$ (contradicting $\overset{\sim 1}{k_p} \leq k_p$);
- or $Pr(a_p = 1 | \theta, \omega = 1, k_c > \overset{\sim 1}{k_c}) = \rho_\theta$, with $\rho_1, \rho_0 > 0$ (and both types playing $a_p = 0$ when $\omega = 0$, or when $\omega = 1$ and $k_c < \overset{\sim 1}{k_c}$). Then A3 requires $\rho_1 = \rho_0$, which implies $\mu^{1y0} \geq \pi$ for $y = 0, 1$, and thus $s = t = \beta$; but $\widehat{k}_p^1(s = t = \beta) > k_p$, contradicting $k_p > \widehat{k}_p^1$.

Among the $k_p \leq \widehat{k}_p^0$ cases:

(iv) $k_p \leq \widehat{k}_p^0$ and $k_p > \overset{\sim 1}{k_p}$: $a_p = 1$ is never reached by L^1 on-path. Either $a_p = 1$ is also not reached by L^0, in which case $\mu^{1y0} = \pi$ by A2, so $s = t = \beta$ and $\overset{\sim 1}{k_p}(s = t = \beta) > k_p$ (contradicting $\overset{\sim 1}{k_p} < k_p$); or $a_p = 1$ is reached by L^0, in which case $\mu^{1y0} = 0$, so $s = t = 0$, and $\widehat{k}_p^0(s = t = 0) \leq \alpha_p^0(1 - \alpha_c) < k_p$ (contradicting $k_p \leq \widehat{k}_p^0$).

(v) $k_p < \widehat{k}_p^0$ and $k_p \leq \overset{\sim 1}{k_p}$: $a = (0,1)$ is not reached on the path of play. Then either:

- $a = (0,0)$ is also not played, putting $a_p = 0$ off-path entirely; in which case $\mu^{0y0} = \pi$ by A2, so $q = v = \beta$, and $\widehat{k}_p^0(q = v = \beta) < k_p$ (contradicting $k_p < \widehat{k}_p^0$);
- or $a = (0,0)$ is reached on path, with

$$P_1(0,y) = g_y(\alpha_0) \left[\tau(1 - \rho_1^1) + (1 - \tau)(1 - \rho_1^0)\right], \qquad g_y(\alpha) = y\alpha + (1-y)(1-\alpha)$$
$$P_0(0,y) = g_y(\alpha_0) \left[\tau(1 - \rho_0^1)(1 - \kappa_1) + (1 - \tau)(1 - \rho_0^0)(1 - \kappa_0)\right]$$
$$\rho_\theta^\omega = Pr(a = (1,0) | \theta, \omega, k_c > k_c^\dagger) = 1 - Pr(a = (0,0) | ...)$$
$$\kappa_1 = F(\ddot{k}_c^1), \qquad \kappa_0 = \begin{cases} F(\widehat{k}_c^0), & k_p \geq \overset{\sim 0}{k_p} \\ F(\ddot{k}_c^0) & otherwise \end{cases}$$

where $k_c^\dagger$ denotes the pertinent k_c threshold for each respective case ($\ddot{k}_c^1$, $\ddot{k}_c^0$, or $\widehat{k}_c^0$). ρ_θ^ω can only be properly mixed at $k_p = \overset{\sim \omega}{k_p}$, and only characterizes mixing probabilities when $k_c > k_c^\dagger$, in which case both leaders' best-response set is $\{(1,0),(0,0)\}$. Applying A3 gives us $\rho_1^\omega = \rho_0^\omega$, so we have $P_1(0,y) \geq P_0(0,y)$ for $y = 0,1$, for any $\kappa_0, \kappa_1 \in [0,1]$. This means $\mu^{0y0} \geq \pi$ and thus $q = v = \beta$; but $\widehat{k}_p^0(q = v = \beta) < k_p$ (contradicting $k_p < \widehat{k}_p^0$).

■

Lemma 12 (Conditions under which only PAPE are supported) *Let $\rho_\theta^\omega = Pr(a_p = 1|\theta, \omega)$. Define a Public Action Pandering Equilibrium (PAPE) as an equilibrium in which $\rho_1^0 > 0$. All equilibria satisfying $k_p \leq \tilde{k}_p^0$ are PAPE.*

Proof of Lemma 12: The lemma is trivially satisfied for $k_p < \tilde{k}_p^0$, given Lemma 1. What remains to show is that for $k_p = \tilde{k}_p^0$, where L^1 is indifferent between $a_p = 0$ and $a_p = 1$ when $\omega = 0$, L^1 must play $a_p = 1$ with positive probability.

Suppose $\tilde{k}_p^0 = k_p$ and L^1 plays $a_p = 0$ when $\omega = 0$ (that is, $\rho_1^0 = 0$, violating the PAPE definition). This means $P_1(0,0) = (1-\tau)(1-\alpha_0)$. When $k_c > k_c^*$ and $\omega = 0$, at $k_p = \tilde{k}_p^0$, L^0's best-response set is $\{(1,0),(0,0)\}$; so by A3, L^0 must also break her indifference in favor of playing $a_p = 0$. From Lemma 11, we know that $k_p < \widehat{k}_p^1$. So $P_0(0,0) \leq (1-\tau)\left[\kappa_0(1-\alpha_c)(1-\lambda) + (1-\kappa_0)(1-\alpha_0)\right]$ for $\kappa_0 = F(k_c^*)$ (with $P_0(0,0)$ decreasing further if $\phi > 0$); and thus $P_0(0,0) \leq P_1(0,0)$. This means the audience must play $v = \beta$, and $\tilde{k}_p^0(v = \beta) < k_p$ for all q, s, t, violating $\tilde{k}_p^0 = k_p$. ■

Lemma 13 (Conditions under which only PIPE are supported) *Define a Public Inaction Pandering Equilibrium (PIPE) as an equilibrium in which $\rho_1^1 = Pr(a_p = 1|\theta = 1, \omega = 1) < 1$. All equilibria satisfying $k_p > \tilde{k}_p^1$, or satisfying $k_p = \tilde{k}_p^1$ and $\phi = 0$, are PIPE.*

Proof of Lemma 13: The lemma is trivially satisfied for $k_p > \tilde{k}_p^1$, given Lemma 1. What remains to show is that for $k_p = \tilde{k}_p^1$, where L^1 is indifferent between $a_p = 0$ and $a_p = 1$ when $\omega = 1$, L^1 must play $a_p = 0$ with positive probability if $\phi = 0$.

Suppose $\tilde{k}_p^1 = k_p$, and L^1 plays $\rho_1^1 = 1$, violating the PIPE definition. At $k_p = \tilde{k}_p^1$, when $k_c > \ddot{k}_c^1$ and $\omega = 1$, L^0's best-response set is $\{(1,0),(0,0)\}$; so by A3, L^0 must also break her indifference in favor of playing $a_p = 1$. By Lemma 11, we know that $k_p \leq \widehat{k}_p^1$. This gives:

$$
\begin{aligned}
&P_1(1,y) = \tau g_y(\alpha_p^1), \qquad g_y(\alpha) = y\alpha + (1-y)(1-\alpha) \\
&P_0(1,y) = \tau\left[\kappa_1 \psi g_y(\alpha_{pc}^1)(1-\lambda) + (1-\kappa_1) g_y(\alpha_p^1)\right] + (1-\tau)\kappa_0 \phi g_y(\alpha_{pc}^0)(1-\lambda) \\
&\qquad \kappa_1 = F(\ddot{k}_c^1), \quad \kappa_0 = F(k_c^*) \\
&\qquad \psi = Pr(a = (1,1)|\theta = 0, \omega = 1, k_c < \ddot{k}_c^1) = 1 - Pr(a = (0,1)|...)
\end{aligned}
$$

The second term of $P_0(1,y)$ drops out, given the hypothesis $\phi = 0$. Then observe that $\mu^{1y0} - \frac{1}{2}$ shares the sign of $\frac{\pi}{1-\pi}P_1(1,y) - P_0(1,y)$, which is decreasing in ψ. Setting $\psi = 1$ yields the same $P_\theta(1,y)$ expressions from the NCRE; and Lemma 8 showed that these $P_\theta(1,y)$ quantities imply $\mu^{100} > \frac{1}{2}$ and $\mu^{110} > \frac{1}{2}$. Thus $s = t = \beta$ and $\tilde{k}_p^1(s = t = \beta) > k_p$ for any q, v (Remark 6), contradicting $\tilde{k}_p^1 = k_p$. ■

Remark 11 (Restatement of Lemma 11) *Lemma 11 can be restated as follows:*
All equilibria fall into one of three categories:

- *a Responsive Equilibrium (RE), with* $\tilde{k}_p^0 < k_p$ *and* $\widehat{k}_p^0 \leq k_p < \widehat{k}_p^1$*, which is either (i) interior, with* $k_p < \tilde{k}_p^1$*, or (ii) a CSRE with* $k_p = \tilde{k}_p^1$*.*[27]
- *a Public Action Pandering Equilibrium (PAPE), with* $k_p \leq \tilde{k}_p^0$ *and* $\widehat{k}_p^0 \leq k_p < \widehat{k}_p^1$*;*
- *or a Public Inaction Pandering Equilibrium (PIPE), with* $k_p \geq \tilde{k}_p^1$ *and* $\widehat{k}_p^0 \leq k_p \leq \widehat{k}_p^1$ *(where* $k_p = \widehat{k}_p^0$ *only if* $k_p = \tilde{k}_p^1$*).*

Proof of Remark 11: The classification of equilibria into RE/PAPE/PIPE is exhaustive, by the definitions of each equilibrium class (each is defined by L^1's behavior, ρ_1^0 and ρ_1^1). They are also mutually exclusive (and thus a partition of the set of possible equilibria):

- RE (which requires $\rho_1^1 = 1, \rho_1^0 = 0$) is disjoint from PAPE and PIPE, by their definitions.
- Following from the PIPE definition: $\rho_1^1 < 1 \implies k_p \geq \tilde{k}_p^1 > \tilde{k}_p^0 \implies \rho_1^0 = 0$, by Lemma 1, so any PIPE fails the PAPE definition. Conversely, because $\rho_1^0 > 0 \implies \rho_1^1 = 1$, any PAPE fails the PIPE definition.

The claims regarding how this partition corresponds with the k_p boundary conditions follow from Lemmas 11, 12, and 13. ■

Proposition 5 (Payoff-dominance of NCRE under high transparency) *If* $\lambda > \bar{\lambda}$*, and* α_0 *is small (below* $\underline{\alpha}_0$*, defined in the proof that follows), the interior NCRE exists and yields a higher ex-ante payoff for* L^1 *than does any other equilibrium.*

Proof of Proposition 5: From Propositions 2 and 3, we know that an interior NCRE exists whenever $\lambda > \bar{\lambda}$. From Remark 11, we know that there are (at most) three other categories of equilibria that we need to compare the interior NCRE against: the PIPE, the PAPE, and the "corner CSRE" at $k_p = \widehat{k}_p^0 = \tilde{k}_p^1$. We will consider the worst-case L^1 payoff from the interior NCRE, and show that it is greater than the best-case payoffs for each of the other options.

Recall that the interior NCRE involves L^1 playing $a_p = \omega, a_c = 0$, with the audience playing $\sigma_A = (s = t = v = \beta, q = q^*(0))$. To put a lower bound on L^1's payoff, suppose $q = 0$. Then:

$$EU_L^1(\text{NCRE}) \geq \tau\left[\alpha_p^1(1+\beta) + (1-\alpha_p^1)\beta - k_p\right] + (1-\tau)\left[\alpha_0 + (1-\alpha_0)\beta\right]$$

Interior NCRE payoff-dominates PAPE (unconditional on α_0): Among PAPE (whether CSE-PAPE or non-CSE-PAPE), L^1's payoff is maximized either when $k_p < \tilde{k}_p^0$, or when $k_p = \tilde{k}_p^0$; but in the latter case, at $\omega = 0$, she is indifferent between $a_p = 0$ vs. $a_p = 1$. Thus her payoff is bounded above by what she achieves from always playing $a_p = 1$, with $\sigma_A = (\beta, \beta, \beta, \beta)$. So:

$$EU_L^1(\text{PAPE}) \leq \tau\left[\alpha_p^1(1+\beta) + (1-\alpha_p^1)\beta - k_p\right] + (1-\tau)\left[\alpha_p^0(1+\beta) + (1-\alpha_p^0)\beta - k_p\right]$$

Comparing $EU_L^1(\text{PAPE})$ against $EU_L^1(\text{NCRE})$, we see that the payoffs can only differ in $\omega = 0$. Thus we want to show:

$$EU_L^1(\text{PAPE}; \omega = 0) \leq \alpha_p^0(1+\beta) + (1-\alpha_p^0)\beta - k_p < \alpha_0 + (1-\alpha_0)\beta \leq EU_L^1(\text{NCRE}; \omega = 0)$$
$$\alpha_p^0 - k_p < \alpha_0(1-\beta)$$

[27] Note that we have not proven existence of a CSRE at $k_p = \tilde{k}_p^1$, but we also have not shown that it can be ruled out. Recall also that any CSE requires $k_p \leq \widehat{k}_p^0$; so this CSRE, if it exists, requires $k_p = \tilde{k}_p^1 = \widehat{k}_p^0$.

LHS $< 0 <$ RHS, so this is always satisfied.

Interior NCRE payoff-dominates PIPE and corner CSRE (conditional on low α_0): We first note that if the corner CSRE exists, we can characterize the same upper-bound L^1 payoff for the corner CSRE as we can for the PIPE. In both cases, suppose the best-case reputational payoff $\sigma_A = (\beta, \beta, \beta, \beta)$ (which, for this bounding exercise, need not be sequentially rational for the audience). Analogously to the PAPE case, we can see that in the PIPE, L^1's maximum payoff can be achieved by always playing $a_p = 0$ (either because she strictly prefers it at $k_p > \tilde{k}{}_p^1$, or because she is indifferent at $k_p = \tilde{k}{}_p^1$ when $\omega = 1$). The same applies to the corner CSRE: at $k_p = \tilde{k}{}_p^1$, L^1 is indifferent between $a_p = 0$ and $a_p = 1$ when $\omega = 1$, which means her highest payoff can be achieved by always playing $a_p = 0$.[28] So:

$$\max \left\{ EU_L^1(\text{PIPE}), EU_L^1(\text{corner CSRE}) \right\} \leq \alpha_0 + \beta$$

Comparing this against the NCRE lower-bound payoff, we want to show

$$\begin{aligned} \alpha_0 + \beta &< \tau \left[\alpha_p^1(1+\beta) + (1-\alpha_p^1)\beta - k_p\right] + (1-\tau)\left[\alpha_0 + (1-\alpha_0)\beta\right] \\ &= \tau \left[\alpha_p^1 - k_p\right] + \tau\beta + (1-\tau)\beta + (1-\tau)\alpha_0 - (1-\tau)\alpha_0\beta \\ \alpha_0\left[\tau + \beta(1-\tau)\right] &< \tau\left[\alpha_p^1 - k_p\right] \end{aligned}$$

Thus we can see that the inequality is satisfied—meaning the interior NCRE payoff-dominates the other two categories of equilibria for L^1—if $\alpha_0 < \underline{\alpha}_0 := \dfrac{\tau(\alpha_p^1 - k_p)}{\tau + \beta(1-\tau)}$, which is positive. ■

Lemma 14 (CSE threshold among PAPE) *In any PAPE, $\phi = 0$ implies $k_p \geq \widehat{k}_p^{PAPE}$, where $\widehat{k}_p^{PAPE} \geq \widehat{k}_p^{NCRE}$.*

Proof of Lemma 14: In a PAPE with $\phi = 0$, audience beliefs in the $a_p = 1$ histories are given by

$$\begin{aligned} &P_1(1,y) = \tau g_y(\alpha_p^1) + (1-\tau)\rho^0 g_y(\alpha_p^0), \qquad g_y(\alpha) = y\alpha + (1-y)(1-\alpha) \\ &P_0(1,y) = m_{pc}^1 g_y(\alpha_{pc}^1) + m_p^1 g_y(\alpha_p^1) + m_p^0 g_y(\alpha_p^0) \\ &\qquad m_{pc}^1 = \tau\kappa_1(1-\lambda), \quad m_p^1 = \tau(1-\kappa_1), \quad m_p^0 = (1-\tau)\rho^0(1-\kappa_0) \\ &\qquad \kappa_0 = F(\tilde{k}{}_c^0(\cdot)), \quad \kappa_1 = F(\ddot{k}_c^1(\cdot)), \quad \rho^0 = Pr(a = (1,0)|\omega = 0, k_c > \tilde{k}{}_c^0) = 1 - Pr(a = (0,0)|...) \end{aligned}$$

where ρ^0 is common across $\theta = 0, 1$ by Assumption 3. (Recall that in any PAPE, we have $k_p < \widehat{k}_p^1$, per Lemma 11.) We can see that $\frac{\pi}{1-\pi}P_1(1,y) - P_0(1,y)$ is increasing in ρ^0 for $y = 0, 1$, and $\rho^0 = 0$ corresponds to the NCRE beliefs. Thus after plugging in $\rho^0 = 0$, the same derivation from Lemma 8 demonstrates that $\mu^{110}, \mu^{100} > \frac{1}{2}$, so $s = t = \beta$.

Thus in a PAPE, $\phi = 0$ requires $k_p \geq \widehat{k}_p^{PAPE} := \widehat{k}_p^0(s = t = \beta, \tilde{q}, \tilde{v})$ for equilibrium values $\tilde{q}, \tilde{v}$. Observe that $\widehat{k}_p^0$ is decreasing in both q and v. We want to show that $\widehat{k}_p^{PAPE} \geq \widehat{k}_p^{NCRE}$. Because the σ_A values that determine $\widehat{k}_p^{NCRE}$ are $\sigma_A = (s = t = v = \beta, q = q^*(\phi = 0))$, and because $\tilde{v} \leq \beta$, it follows that $\tilde{q} \leq q^*(0)$ is a sufficient condition for $\widehat{k}_p^{PAPE} \geq \widehat{k}_p^{NCRE}$.

If $k_p \leq \tilde{k}{}_p^0$ and $\rho^0 = 1$ (and note that $k_p < \tilde{k}{}_p^0$ requires $\rho^0 = 1$): then $a_p = 0$ is never reached on-path by L^1, so $\tilde{q} = \tilde{v} = 0$, and the proof is complete. (If $a_p = 0$ is also never played by L^0, then

[28] By A3, L^1 must break her indifference in favor of $a_p = 1$ in a corner CSRE; but what matters for present purposes is that her expected $a_p = 1$ payoff is equal to her expected $a_p = 0$ payoff.

$q = v = \beta$ by A2, implying $k_p > \overset{\sim 0}{k}_p$ (Remark 6), a contradiction.)

If $k_p = \overset{\sim 0}{k}_p$ and $\rho^0 < 1$, the audience belief in the $(0,1,0)$ history of the PAPE (when $\phi = 0$) is given by

$$\mu^{010}(\tilde{q}, \tilde{v}) = \frac{\pi\alpha_0}{\pi\alpha_0 + (1-\pi)\left(\kappa_0\left[\frac{\alpha_c(1-\lambda)}{1-\rho^0} - \alpha_0\right] + \alpha_0\right)}, \qquad \kappa_0 = F(\overset{\sim 0}{k}_c(\tilde{q}, \tilde{v})) = F(k_c^*(\tilde{q}, \tilde{v}))$$

with k_c^* increasing in $\tilde{q}$ and decreasing in $\tilde{v}$. (Recall that the thresholds $\overset{\sim 0}{k}_c$ and k_c^* coincide at $k_p = \overset{\sim 0}{k}_p$.) Analogously to Lemma 7, $\tilde{q}$ is uniquely determined as the fixed point to $\mu^{010}(\tilde{q}, \tilde{v}) = \frac{1}{2}$ when it exists, or a boundary otherwise (0 when $\mu^{010}(0, \tilde{v}) < \frac{1}{2}$, or β when $\mu^{010}(\beta, \tilde{v}) > \frac{1}{2}$). For comparison, $q^*(0)$ is the corresponding value when $\tilde{v} = \beta$ and $\rho^0 = 0$.

We want to show that the fixed point $\tilde{q}$ is no greater than the fixed point $q^*(0)$. We begin by noting two partial derivatives of μ^{010}: first, μ^{010} is decreasing in ρ^0, which enters μ^{010} only via the $\frac{\alpha_c(1-\lambda)}{1-\rho^0}$ in the denominator; second, μ^{010} is increasing in $\tilde{v}$, which only enters via $\kappa_0 = F(k_c^*)$ in the denominator, with k_c^* decreasing in $\tilde{v}$. Note also that the PAPE, by definition, requires $\rho^0 > 0$. Thus for any fixed q, and any $\tilde{v} \leq \beta, \rho^0 > 0$, we have

$$\mu^{PAPE}(q) := \mu^{010}(q; \tilde{v}, \rho^0, \phi = 0) < \mu^{010}(q; v = \beta, \rho^0 = 0, \phi = 0) =: \mu^{NCRE}(q)$$

So $\mu^{PAPE} < \mu^{NCRE}$ pointwise in q, and both are decreasing in q (which only enters via $\kappa_0 = F(k_c^*)$ in the denominator, with k_c^* increasing in q).

Then we have two cases to consider:

- If $\tilde{q} > 0$, then $\mu^{PAPE}(\tilde{q}) \geq \frac{1}{2}$, so $\mu^{NCRE}(q) > \frac{1}{2}$ for any $q < \tilde{q}$; but no such q can be the NCRE equilibrium value, as the belief $\mu^{NCRE}(q) > \frac{1}{2}$ requires $q = \beta$. Therefore, the fixed-point in the NCRE requires $q^*(0) \geq \tilde{q}$.

- If $\tilde{q} = 0$, then trivially, $q^*(0) \geq \tilde{q}$.

In both cases, we have $\tilde{q} \leq q^*(0)$. Therefore, $\widehat{k}_p^{PAPE} \geq \widehat{k}_p^{NCRE}$. ■

Proposition 6 (Payoff-dominance of CSE under low transparency) *If $\lambda < \bar{\lambda}$:*

- *All equilibria are either CSE or PIPE (or both).*
- *If α_0 is small (below $\underline{\underline{\alpha}}_0$, defined in the proof that follows), there exists a CSRE that yields a higher ex-ante payoff for L^1 than does any non-CSE PIPE.*

Proof of Proposition 6: For the first claim of the proposition, it suffices to show that if $\lambda < \bar{\lambda}$, then (i) all interior RE are CSE, and (ii) all PAPE are CSE. (i) follows directly from Proposition 3. (ii) follows from the fact that $\lambda < \bar{\lambda} \implies k_p < \widehat{k}_p^{NCRE} \leq \widehat{k}_p^{PAPE}$: Lemma 14 gives us both the inequality $\widehat{k}_p^{NCRE} \leq \widehat{k}_p^{PAPE}$, and (by contraposition) that within PAPE, $k_p < \widehat{k}_p^{PAPE} \implies \phi > 0$.

For the second claim, we will compare L^1's ex-ante payoffs from the interior CSRE characterized in Lemma 9 against her best-case ex-ante payoffs from a non-CSE PIPE.

Consider audience beliefs in the non-CSE PIPE (given $\phi = 0$):

$$
\begin{aligned}
&P_1(0,1) = \tau(1-\rho^1)\alpha_0 + (1-\tau)\alpha_0 \\
&P_0(0,1) = \tau\left[\kappa_1(1-\psi)\alpha_c(1-\lambda) + (1-\kappa_1)(1-\rho^1)\alpha_0\right] + (1-\tau)(\kappa_0\alpha_c(1-\lambda) + (1-\kappa_0)\alpha_0) \\
&\qquad \kappa_0 = F(k_c^*(q,v)), \quad \kappa_1 = F(\widehat{k}_c^1(\cdot)) \\
&\qquad \rho^1 = Pr(a=(1,0)|\omega=1, k_c > \widehat{k}_c^1) = 1 - Pr(a=(0,0)|...) \\
&\qquad \psi = Pr(a=(1,1)|\theta=0, \omega=1, k_c < \widehat{k}_c^1) = 1 - Pr(a=(0,1)|...)
\end{aligned}
$$

(where ρ^1 is constant across $\theta = 0, 1$ by A3). Recall that $q > 0$ requires $\mu^{010} \geq \frac{1}{2}$, or $\frac{\pi}{1-\pi}P_1(0,1) \geq P_0(0,1)$. Observe that $P_1(0,1) \leq \alpha_0$ for any $\rho^1 \in [0,1]$; and $P_0(0,1) \geq (1-\tau)\tilde{\kappa}_0\,\alpha_c(1-\lambda)$ for any $\rho^1, \kappa_1, \psi \in [0,1]$ and any $q, v \in [0,\beta]$, where $\tilde{\kappa}_0 := F(\min_{\sigma_A} k_c^*(\sigma_A)) = F(k_c^*(q=0, v=\beta)) > 0$ by A1(vi).[29]

Thus $\mu^{010} \geq \frac{1}{2}$ requires $\frac{\pi}{1-\pi}\alpha_0 \geq (1-\tau)\tilde{\kappa}_0\alpha_c(1-\lambda)$. The LHS is increasing in α_0, while RHS is non-increasing in α_0 (since $k_c^*(0,\beta) = \alpha_c - \alpha_0 + \beta[(1-\alpha_c)(1-\lambda) - (1-\alpha_0)]$). Thus there exists some $\underline{\underline{\alpha}}_0 > 0$ such that $\alpha_0 < \underline{\underline{\alpha}}_0 \implies \mu^{010} < \frac{1}{2} \implies q = 0$ in any non-CSE PIPE. We assume $\alpha_0 < \underline{\underline{\alpha}}_0$ going forward.

L^1's ex-ante expected payoff from the non-CSE PIPE is maximized at either $k_p = \overset{\sim 1}{k_p}$ or $k_p > \overset{\sim 1}{k_p}$. If $k_p = \overset{\sim 1}{k_p}$, then she is indifferent between her payoffs when playing $a = (1,0)$ vs. $a = (0,0)$ when $\omega = 1$: holding fixed σ_A, if some $\rho^1 > 0$ gives L^1 her highest possible payoff across PIPE, then that same payoff is achieved by $\rho^1 = 0$. Thus for simplicity we focus on her payoff from always playing $a = (0,0)$, with the best-case audience reward $v = \beta$; this serves to characterize her best-case ex-ante expected payoff across all non-CSE PIPE (with $k_p = \overset{\sim 1}{k_p}$ or $k_p > \overset{\sim 1}{k_p}$). Thus we have

$$EU_L^1(PIPE) \leq \alpha_0(1+q) + (1-\alpha_0)\beta = \alpha_0 + (1-\alpha_0)\beta$$

In the interior CSRE characterized in Lemma 9, her ex-ante expected payoffs are given by

$$EU_L^1(CSE) = \tau\left[\alpha_p^1(1+s^*(\phi^*)) + (1-\alpha_p^1)\beta - k_p\right] + (1-\tau)\left[\alpha_0(1+q^*(\phi^*)) + (1-\alpha_0)\beta\right]$$

with a lower bound achieved by plugging in $q^*(\phi) = 0$ (both in the displayed equation and in the expression for $s^*(\phi^*)$, since s^* is increasing in q, as shown in Lemma 9). Thus we want to show:

$$
\begin{aligned}
\alpha_0 + (1-\alpha_0)\beta &< \tau\left[\alpha_p^1(1+s^*|_{q=0}) + (1-\alpha_p^1)\beta - k_p\right] + (1-\tau)\left[\alpha_0 + (1-\alpha_0)\beta\right] \\
0 &< \left[\alpha_p^1(1+s^*|_{q=0}) + (1-\alpha_p^1)\beta - k_p\right] - \left[\alpha_0 + (1-\alpha_0)\beta\right] \\
&= (\alpha_p^1 - \alpha_0)(1-\beta) - k_p + \frac{\alpha_p^1}{\alpha_{pc}^0}\left[\frac{k_p - \alpha_p^0(1-\alpha_c)}{1-\lambda} + \alpha_p^0(1-\alpha_c)\beta\right]
\end{aligned}
$$

Observe that this is (12) from the proof of Lemma 9 (evaluated at $q^* = 0$)—which is implied by (13), a sufficient condition for non-deviation from the interior CSRE (which was shown to be satisfied by A1(i)). The reason for this correspondence is intuitive: by deviating from $a = (1,0)$ to $(0,0)$ when $\omega = 1$, L^1 can guarantee herself the PIPE payoff of $\alpha_0 + (1-\alpha_0)\beta$ (given $\alpha_0 < \underline{\alpha}_0 \implies q = 0$). ■

[29] As in the proof of Corollary 1, hold fixed $(\underline{k_c}, \overline{k_c})$ and all other primitives as α_0 varies.