Cover Stories

Michael F. Joseph
UCSD
mfjoseph@ucsd.edu

Matt Malis
Texas A&M University
malis@tamu.edu

February 6, 2026

Abstract

How do governments maintain plausible deniability for their controversial covert actions? While existing research focuses on the risk of *direct* exposure, we contribute by highlighting the challenges posed by *circumstantial* evidence, and the inferences that audiences can draw from their knowledge of the strategic environment. Through a formal model, we uncover a novel "cover story" mechanism, whereby governments use ineffective public action alongside effective covert action, to provide an alternative explanation for how a policy outcome came about. We illustrate this mechanism through detailed examination of the CIA's Operation PBSUCCESS (Guatemala, 1954), along with additional case evidence from treaty negotiations between Australia and East Timor, and the resolution of the Cuban Missile Crisis.

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 (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 the only factor 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 only rely 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 different means: a public action that an audience directly observes, and a secret action that is only observed probabilistically. 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.

The government's optimal strategy depends on *transparency*—the probability that direct evidence of covert action is exposed. Under high transparency, the government is unlikely to use covert action, and the audience does not suspect that a policy success is attributable to covert action. When transparency is low, however, the audience is no longer willing to give the government the "benefit of the doubt". Rather, if they observe a successful outcome despite public inaction, they will infer that covert action was taken out of public view, and punish the government just as if direct evidence had been exposed. To overcome this problem, the government uses performative overt action as a cover story. The cover story succeeds if it mitigates suspicion enough for the government to avoid backlash from the audience.

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, 2021)—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" (Gleason, 2017), admitting ignorance (Backus and Little, 2020), and adopting extreme ideological stances (Izzo, 2022). 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 and costly policies.

1 Secrecy and Plausible Deniability

Policymakers are often tempted to use secrecy to achieve a policy success while avoiding backlash. For example, the US public generally wants to control immigration at the southern border, but they do not want the government to achieve this goal by locking children in cages. The first Trump administration initially sought to conceal its policy of family separation from the US and foreign publics, and succeeded in keeping the policy secret for several months (Horowitz, 2021). The public broadly wants the government to make scientific advancements, but not at the expense of unethical research practices. Government scientists during the 1950s–1970s thus administered unethical experiments on remote, marginalized communities—ethnic minorities, prisoners, and the mentally ill—hoping to reap the policy benefits while concealing the controversial research practices that contributed to breakthroughs (ACHRE, 1996).

While the policy domain is potentially 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 an audience that generally deems the covert action to be "unscrupulous", and in conflict with principles that the audience values. The Intervener worries that an exposed covert action will tarnish their reputation, resulting in some form of punishment from the audience.

Previous studies have considered audiences in the form of an international community who cares if the Intervener violates international laws and norms (Poznansky, 2025; Bull, 2002); Congress or other political elites who want the president to adhere to institutional constraints (Smith, 2019; Colaresi, 2012); or the press and the broader public, who care about principles of transparency and integrity in government (Spaniel and Poznansky, 2018; Myrick, 2020). Our abstract theoretical model can incorporate this variety of substantive considerations.

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 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 the possibility of reputational costs arising from inference or speculation on the part of the audience.

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 begin with a general definition:

Definition 1 (Cover story) *A cover story is a publicly observed government action which reduces an audience's belief that the government implemented a distinct, unobserved, secret action to achieve a recently observed policy success.*

We now describe a simple model of audience learning about government policymaking, to identify a set of basic conditions under which a cover story can arise. The model takes the government's behavior as given, and asks what the audience can infer about unobserved behavior given the information they observe. In particular, 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 provided a cover story.

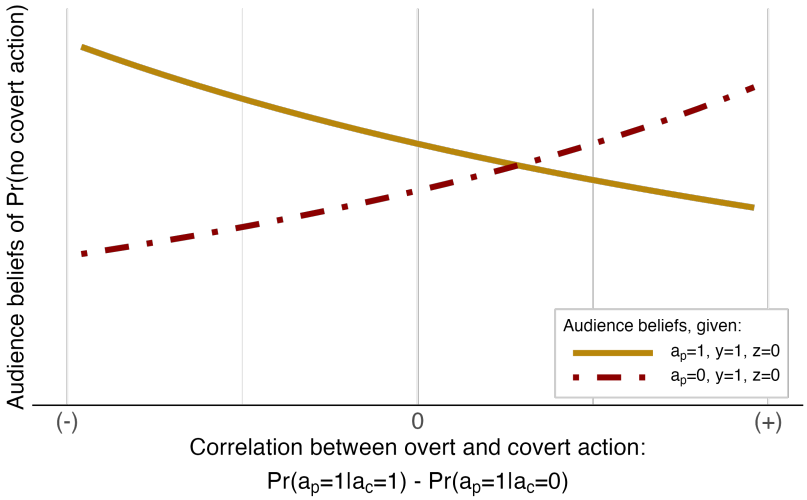
The model's essential features are as follows (Appendix 7.1 presents technical details). The government takes either public action ($a_p \in \{0,1\}$), covert action ($a_c \in \{0,1\}$), both, or neither, according to a commonly known probability distribution. The use of public and covert action can be positively or negatively correlated. The audience observes the public action directly. 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 actions can be complements or substitutes in the policy production function.

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)).

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.*

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.

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. To understand when and why this occurs, first consider the audience's beliefs when covert and public action are uncorrelated (horizontally in the center of Figure 1). In this case, if 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.

The gap between the two audience beliefs widens as we move leftward from the center of Figure 1, reflecting both the indirect reasoning just discussed, and the fact that the use of public action directly implies that covert action is less likely to have been used. Notably, the public action still serves as a cover story even when public and covert action are somewhat positively correlated. This remains true as long as the *indirect* reasoning regarding the need to explain a successful outcome dominates the *direct* reasoning regarding the leader's correlated use of both policy levers.

These results show how cover stories can emerge under general conditions. The analysis so far has been entirely agnostic as to *why* the government is taking these actions with the specified probabilities, or what the government's objective function is. In reality, we should expect the government to strategically select policies, conscious of both the desired outcome and the audience's beliefs regarding the means by which it was achieved. Such strategic government behavior will in turn alter the audience's beliefs regarding the government's actions, creating a complex feedback process. Consequently, it is unclear whether and under what conditions a cover story can be deliberately used by the government for reputational management, and still influence audience beliefs as the government intends. We address these questions next, through game-theoretic analysis.

3 A Strategic Model of Covert Action and Cover Stories

3.1 Model Setup

We study an 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.[1] 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 slashing sugar quotas and oil embargoes, while a_c can represent CIA efforts to assassinate Castro or oust him via support of Cuban exiles.

[1] Section 7.5 explains how our core results obtain in a model with homogeneous leader types.

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 $k_c \in [\underline{k_c}, \overline{k_c}]$, 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

Before L decides which policies to authorize, in step 2, L receives private information about 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 support $[\underline{k_c}, \overline{k_c}]$, 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, as we discuss below. The leader observes k_c directly, while the audience only knows its distribution.[2]

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

[2] Assuming that k_c is privately observed is not necessary for the model's results; see Appendix 7.2.

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 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 $\alpha_0 \leq \alpha_p^0 < \alpha_p^1 < 1$: 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 $\alpha_0 < \alpha_c < 1$ (but impose no restriction on the relative efficacy of covert vs. public action). 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

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

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

[4]The two policy levers are complements $\iff \alpha_{pc}^{\omega} - \alpha_p^{\omega} > \alpha_c - \alpha_0 \iff \alpha_0 > \alpha_p^{\omega}\alpha_c$.

Whenever L refrains from covert action, A observes $z = 0$;[5] 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.

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:[6]

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

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. For a legislature, it can represent the choice to punish the executive by withholding funding of its policy priorities, curtailing its statutory authority, or imposing political costs through public investigations (Colaresi, 2012; Spaniel and Poznansky, 2018). 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 (Poznansky, 2025). In each case, the audience prefers to "reward" the leader if and only if the leader is scrupulous, but is uncertain about the leader's true type. We assume $\pi > \frac{1}{2}$, meaning that 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 β 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

[5] This means that A never observes "false positives" (i.e. *direct* evidence of covert action that was not taken). See Appendix 7.2 for discussion.

[6] See Appendix 7.2 for discussion of alternative assumptions regarding audience preferences.

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 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,[7]

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

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.

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. 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

[7] This parameter restriction is imposed by Assumption 1.

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 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.

Proposition 1 *There exist thresholds λ^* and λ^{**} such that:*[8]

- *If transparency is high ($\lambda \geq \lambda^{**}$), the leader never takes covert action.*
- *If transparency is at an intermediate level ($\lambda^* < \lambda < \lambda^{**}$), the leader takes covert action with positive probability, but never uses a cover story.*
- *If transparency is low ($\lambda < \lambda^*$), the leader uses a cover story with positive probability; that is, all equilibria are CSE.*

The following sections discuss the intuition behind the equilibrium behavior we observe across these ranges of transparency. Proofs for all formal results are presented in the appendix.

3.2.1 The disciplining effect of 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

[8] This result invokes assumptions regarding parameter restrictions and equilibrium selection; see Appendix 7.2.

scrupulousness ($\mu^{a_p,y,z=1} = 0$), and punishes the leader accordingly. When transparency is high ($\lambda \geq \lambda^{**}$) the risk of direct exposure and ensuing punishment disciplines the unscrupulous leader into never taking covert action.[9] Under this condition, both leader-types exhibit the same behavior: they take only public action given a favorable state ($a = (a_p = 1, a_c = 0)$ when $\omega = 1$), and otherwise do nothing ($a = (0, 0)$ when $\omega = 0$).

Now suppose that, despite the leader's inaction, the audience observes an "unexplained success"—that is, a successful outcome with no public action and no direct evidence of covert action (formally, $h = (a_p = 0, y = 1, z = 0)$). Knowing that the leader *never* takes covert action under high transparency, the audience rationally attributes the unexplained success to random luck (which can occur with probability α_0). Consequently, their belief of the leader's scrupulousness remains unchanged ($\mu^{a_p=0,y=1,z=0} = \pi$), and the audience does not punish the leader (because $\pi > \frac{1}{2}$).

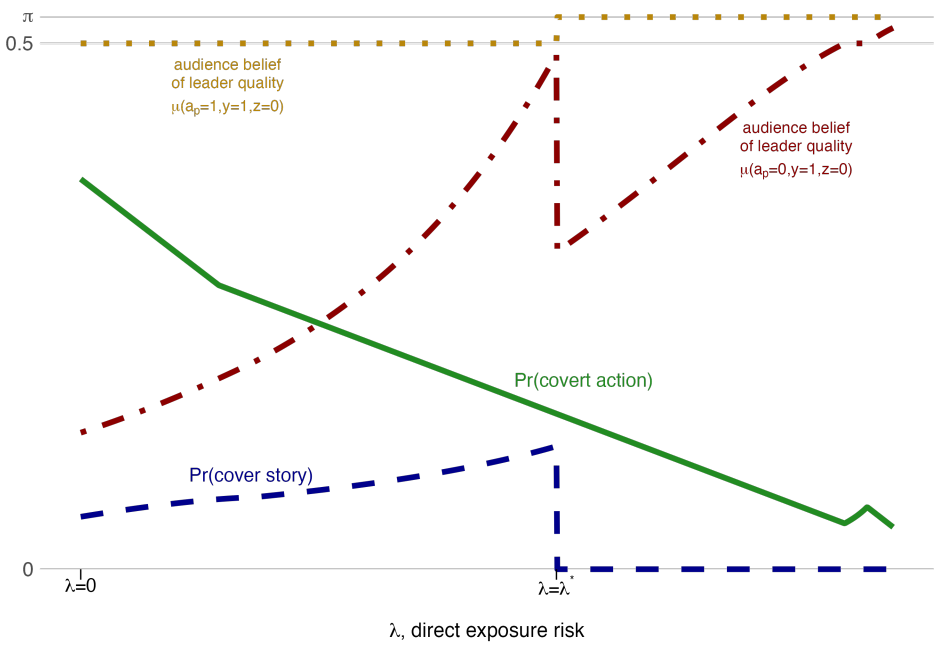
3.2.2 The problem of unexplained success

When transparency falls into the intermediate range (λ falls below λ^{**}), covert action (sometimes) becomes a worthwhile gamble for the leader. If the direct costs (k_c) are low, the leader will accept the risk of exposure in exchange for increasing her chances of a policy success. This complicates the audience's inference. 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; 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. As transparency decreases, the leader becomes more likely to take covert action, and the audience becomes less likely to observe covert action if it is taken. Both contribute to audience suspicion following an unexplained success.

This shift in leader strategy and audience beliefs is visualized in Figure 3. On the far righthand side (highest transparency), the leader uses covert action only infrequently; after observing an unexplained success, the audience is willing to believe that the outcome is attributable to random luck, and refrains from punishing the leader. This favorable inference can only be sustained up to a point, however. As transparency decreases, the absence of direct evidence of covert action becomes less informative as to whether or not covert action was actually taken, and the audience's belief

[9] Note that, under some parameter values, λ^{**} may be above 1.

Figure 3: Leader Strategy and Audience Beliefs



Note: Figure depicts leader strategies and audience beliefs in an equilibrium characterized by Proposition 1. π is the prior probability the leader is scrupulous; $\mu(a_p, y, z)$ denotes the audience's posterior belief. Note that under the parameter values used to construct this figure, $\lambda^{**} > 1$, so the figure only depicts equilibrium behavior within the "low" and "intermediate" ranges of transparency.

places more weight on covert action rather than random luck being the cause of an unexplained success. Eventually, the audience's posterior belief $\mu^{a_p=0,y=1,z=0}$ (the red dot-dashed line in the figure) falls below $\frac{1}{2}$, and they fully punish the leader purely on the basis of circumstantial evidence.

At the same time, in the intermediate range of transparency ($\lambda^* < \lambda < \lambda^{**}$), the audience's belief is more favorable after seeing the leader take public action (because unscrupulous leaders are no more likely than scrupulous leaders to take public action). This is visualized in the gold dotted line in the figure remaining flat at π over the range of $\lambda > \lambda^*$. 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 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 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 simultaneously 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.[10]

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 when transparency is low, $\lambda \leq \lambda^*$. The threshold λ^* 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, λ^* 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*

[10]For technical details, see Remarks 6 and 7 in Appendix 7.4.

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 α_c), since the audience has little reason to suspect 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, and chooses whether to punish or reward the leader given all three pieces of information. 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 any 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 a 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, relevant to all historical research into covert action, highlights strategic complications that mission planners must navigate to achieve plausible deniability:

Implication 1. In any observed case of covert action, covert mission planners within Intervener states should not only be concerned with operational security and the risk of direct exposure of their actions; 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, in cases where covert mission planners are *least* concerned with the risk of direct exposure, they should be *most* concerned about being blamed by the audience in the event of an unexplained success.

In light of these concerns, when Interveners use covert action under conditions of low transparency, we further 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 that the policymaker took which ostensibly contributed to the policy success.

3.4 Alternative Theories of Joint Covert and Overt Actions

We hope to encourage further research on the strategic interdependence between open and secret government actions across a range of political contexts. To aid in this effort, we briefly discuss two potential alternative explanations for the joint use of secret and overt actions: a "policy buildup" and "comprehensive pressure".

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. One key distinction from a cover story is that under a policy buildup, the leader originally hopes to achieve the desired 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 the public action, or the decisions to authorize both actions were made 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. There are many variants of comprehensive pressure, but two seem especially relevant in the international security context. First, leaders may want to maximize their chances of achieving a primary policy objective (for instance, removing a foreign leader or regime from power), and view both covert and public actions as substantially increasing the probability of success. Second, leaders 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 to be counterproductive, or too ineffective to justify the associated costs. Second, comprehensive pressure cannot explain why overt actions would be pursued in ways that draw unnecessary attention to government policies without enhancing the chance of success. Third, cover stories and comprehensive pressure may imply divergent impacts of the public action on audience beliefs. Under some variants of comprehensive pressure described above, if it were the primary motivation for overt action, we would expect a positive correlation between the uses of covert and public action. 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.

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."[11]

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, 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.

[11] Quoted in Schmitz (1999, 179)

4.1 Case Selection, and Calibrating the Parameters

Following Bates (1998), we identified Eisenhower's Guatemala intervention as a case fitting the core strategic tension characterized by our model. Eisenhower[12] believed that stopping the spread of communism required that the US maintain a reputation as promoting 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), Poznansky (2019, 86)).

The conditions of this case fit the parameters of our model that support a cover story equilibrium, consistent with 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 a low level of transparency (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 the fact that Eisenhower was facing reelection at home, 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. In practice, it is difficult for researchers to precisely quantify relative costs and benefits associated with the theoretically optimal combination of overt actions. However, we can gain leverage on this condition by examining policymakers' evaluations of their available policy options.

At multiple points in the years preceding PBSUCCESS, the US government considered various forms of overt action but ultimately decided against them. Truman considered implementing economic sanctions and diplomatic isolationist policies toward Guatemala 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, 1979, 218). A robust policy discussion among State

[12]Given historical debates about Eisenhower's role in foreign policy, and the degree of CIA autonomy, we focus on Eisenhower and the Dulles brothers, as well as their subordinates at CIA and State (Divine, 1981; McAuliffe, 1981).

Department officials about exploiting economic and diplomatic leverage emerged in 1952. But policymakers discounted the possibility, believing that the Soviets were likely meddling covertly, and "it was necessary to fight fire with fire" (Immerman, 1979, 239): some form of US-backed military intervention was necessary to oust Arbenz or alter his policy.[13] 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 deemed not to be viable on their own (Cullather, 2006, 35).

The decision-making processes that led policymakers to dismiss overt actions alone are complex. But one critical factor was that there were no local actors were willing or 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 there is only so much that careful planning can do. With officers stationed across Latin America to train and supply the coup plotters—even opening an operation center inside of Guatemala in December 1953 (Cullather, 2006, App. A)—there always remained a risk of direct exposure, par-

[13]The CIA launched a preliminary covert mission called PBFORTUNE to assassinate Arbenz' political supporters (not overthrow the regime). It is unclear if Eisenhower approved it. PBFORTUNE was greatly scaled back during implementation, and was not exposed.

ticularly after the active phase of PBSUCCESS was given the "full green light" in April 1954.

Given the intense focus on maintaining secrecy, we might expect that the administration would try to divert public attention away from Guatemala as CIA officers were in the field, to minimize the risk of direct 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, Eisenhower forced an anticommunist resolution designed to isolate Guatemala first on the meeting's agenda (Immerman, 1982, ch 19).

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.

Around the same time, the US convened an emergency meeting of the OAS, where Dulles delivered an impassioned speech attacking the Guatemalan government. This was at Eisenhower's direction, 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).

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 that Eisenhower simply 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 requrie 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!"[14]

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-commmunist 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 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

Section 3.3 outlined four novel implications of our strategic theory. We now present empirical support for each.

[14] Quoted in (Taylor, 1956, 794)

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 first establish that policymakers worried ex-ante about audience attribution of covert action in the absence of direct evidence. We find that while planning PBSUCCESS, administration officials expressed acute concerns for strategic inferences. An NSC 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' 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 may never have the opportunity (Cullather, 2006, p42). However, Holland argued for delay because recent factors had raised the risk that US actions 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,[15] 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 position on how best to manage this problem, stating that one view was that this 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 U.S 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-

[15] Only summary minutes from these meetings are available, rather than complete transcripts detailing the participants' logic.

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 as defined in 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.*[16]

[16] CIA (1954*a*); emphasis added.

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.[17]

4.3.4 Implication 4: Audience reception of the cover story

Finally, we expect that observers will naturally be 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.

Several journalists and academics at the time analyzed US involvement in the Guatemalan affair. Notably, 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 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

[17] See CIA (1954*d*) for another similar episode.

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 the logic of 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' favor. In particular, Taylor's interpretation is that US overt actions and public statements contributed to Arbenz' downfall by emboldening Arbenz' opponents to mobilize against him; 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 directly 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.

Perhaps surprisingly, we observe similar reactions from the Arbenz government itself. In a desperate cable to the UNSC, while Armas' invasion was underway, the Guatemalan Foreign Minister denounces a range of US overt actions: he accuses 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 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 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 its publicly observable support for

the overthrow of a democratically elected government. The British Labor Party leader, for instance, expressed that he was "rather shocked at the joy and approval of the American Secretary of State on the success of this *putsch*".[18] Several Latin American governments viewed the blockade as an unjustified violation of sovereignty; but the backlash was relatively minor (Friedman, 2010, 672), compared to the backlash that would have followed from the exposure of PBSUCCESS.

Second, we theorize that cover stories do not conclusively convince the target audience that covert action was not taken. 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 for Overt and Covert Action

We now consider potential confounds, given the two other reasons for joint overt and covert actions detailed in Section 3.4.

Policy buildups involve deploying overt actions before covert actions. As described above, Immerman (1979) 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 (we also discuss Cullather, 2006, 35). Rabe (1988, 43) agrees, noting also that US only started to implement overt policies in late 1953 when "the Department of State began a campaign to convince the other nineteen Latin American nations of the threat that Guatemala posed to their security. *Simultaneously*, President Eisenhower and his most trusted aides secretly developed a plan to destabilize and overthrow Jacobo Arbenz" (emphasis ours). Indeed, the actual coup was implemented long before the US government was able to assess whether any concurrent overt policies would prove to be effective. 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

[18] Quoted in Taylor (1956, 804)

direct effectiveness.

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. The first variant suggests that Eisenhower sought diplomatic isolation to contain Arbenz should PBSUCCESS fail to oust him. Second, Eisenhower could have engaged others at 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 that the suspicions of prominent audiences were mitigated, rather than exacerbated, by observed overt actions.

As a more general takeaway, it is worth emphasizing the following: 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; the alternative theories that we have considered do not.

5 Broadening the Argument

The preceding empirical analysis focused on foreign intervention, a common setting in which scholars have studied questions of culpability for controversial covert action. We now consider two additional vignettes, which illustrate our theory's applicability across different empirical contexts. We summarize these differences in Table 1. While not claiming to provide 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.

	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, 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

5.1 The Timor Gap Scandal

When East Timor seceded from Indonesia in 2002, 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 (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 high-paying jobs.[19] Further, Australia was concerned with other neighboring countries feeling "aggrieved" regarding the status of their own borders with Australia if the border with East Timor became open to reconsideration (Pugh, 2000).

On the other hand, the international community, especially the US, supported more equitable terms. US ambassador Peter Galbraith was appointed to negotiated on behalf of East Timor. As negotiations were ongoing, over 50 US members of congress, including Nancy Pelosi and Patrick Leahy, wrote the Australian Prime Minister calling on Australia to adhere to strict legal principles during the negotiations (Frank, 2004). These American elites, who had supported East Timor's independence at the cost of tension with Indonesia, worried that absent the Timor Gap revenue, the highly oil-dependent East Timor might devolve into a failed state. Another issue for Australia

[19]It was so important that the foreign minister invoked a "national interest" exemption to fast-track ratification of the treaty—an exemption which had only been used six times in Australia's history (Dodd, 2007).

was that East Timor could terminate the existing mining concessions if they did not perceive the agreement as fair, and even escalate the issue to a formal dispute. Legal analysts believed that East Timor could accrue substantial concessions if the matter was referred to 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 Sunrise Gap, and a commitment from East Timor that they could not renegotiate for 30 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 value. Australia attributed their success to an intensive bargaining effort. The government employed expensive outside legal consultants. They withdrew from UNCLOS, an international treaty with broader implications, a month 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, Galbraith and 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, who were invited into East Timor as part of a counter-terrorism operation, had illegally bugged the offices of East Timor's president and other key negotiators (Cannane, 2015). Thus, unbeknownst to Galbraith and the East Timorese, the Australian Government knew exactly what the negotiators were willing to accept. Australia's secret efforts were publicized by a whistleblower, Citizen K, who came forward once he learned that former Australian Foreign Minister Alexander Downer was appointed to the board of Woodside Petroleum, the firm that profited the most from the episode.

The reaction of U.S. elites is consistent with our characterization of audiences who are concerned with preventing unscrupulous policy actions, independent of the ends. According to a journalistic account by Knaus (2019): "As a former US ambassador to Croatia, Galbraith had frequent access to US intelligence. Never has he 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."

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 that was intended to disclaim responsibility for the covert actions which ultimately brought about the desired outcome.

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 crisis' resolution 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[20] 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). During the crisis, Khrushchev and others had raised the exchange, arguing that the proximity of US missiles to the Soviet Union justified the Soviet missile deployment to Cuba. Even at the time, many speculated that a missile exchange could have facilitated peace.[21]

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

[20] Major General William Senter (1963), for instance, expressed concern that Turkey would doubt the US security guarantee if they discovered the US had exchanged the Jupiter missiles.

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

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 sole 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). The non-invasion pledge did not completely offset suspicion. But the amount of suspicion raised did not prevent Democrats from gaining Congressional seats in the following two elections (and maintaining the presidency); nor did it hinder Robert Kennedy's political career, as he feared the Jupiter exchange's public disclosure would.

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 Kennedy's public statements of resolve, he secretly offered the Soviets a substantial concession. This raises the question: why did Kennedy make these public statements at all?

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

We highlight under-appreciated challenges governments face in sustaining plausible deniability given circumstantial evidence. When the risk of direct exposure of covert action is low, strategic

audiences are especially suspicious that a policy success is explained by a controversial secret policy. We argue that cover stories can offset this suspicion, and that forward-looking governments rationally implement overt policies for this purpose.

Our theory holds implications for global security given the re-emergence of great power competition and threats to the Liberal Order. The logic of strategic inferences suggests a practical limitation on how frequently states can use covert action to circumvent international responsibilities without arousing suspicion, especially if they use covert actions repeatedly over time. This complicates existing predictions about grey zone conflict in US relations with Russia and China. Some emphasize that states can avoid escalation and retaliation if they conceal direct evidence of an attack (Carson, 2016; Cormac and Aldrich, 2018; Napier, 2023; Bloch and McManus, 2024). But to the extent that audiences draw strategic inferences, normatively questionable grey zone attacks may not be viable for the US without damaging the Liberal International Order. Our theory shows that cover stories can partly offset this concern.

Our theory also holds implications for domestic politics given growing mistrust in government. For public accountability activists, it explains that building extensive monitoring capabilities may, in some cases, work against their objectives. If the public widely believes that these organizations and the media can effectively scrutinize the government most of the time, then the public will infer from an absence of evidence that no unscrupulous policy took place. It also suggests that policymakers can offset widespread conspiratorial beliefs through performative overt policies. Conspiracies often enter public consciousness when there is no plausible explanation for events that the audience knows to be in the government's interest (Douglas, Sutton and Cichocka, 2017). Cover stories can fill that void.

References

ACHRE. 1996. "Final Report (061-000-00849-7).".

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.

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

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.

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

Bates, Robert H. 1998. *Analytic narratives*. Princeton 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.

Bull, Hedley. 2002. *The anarchical society : a study of order in world politics*. Columbia University Press.

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.

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. 2016. "Facing off and saving face: covert intervention and escalation management in the Korean War." *International Organization* 70(1):103–131.

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

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 Charge d'Affairs a.i. Jan. 25, 1954. (Disbatch 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*.

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.

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.

Cormac, Rory and Richard J Aldrich. 2018. "Grey is the new black: covert action and implausible deniability." *International affairs* 94(3):477–494.

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.

Divine, R A. 1981. *Eisenhower and the Cold War*. Oxford University Press.

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.

Douglas, Karen M, Robbie M Sutton and Aleksandra Cichocka. 2017. "The psychology of conspiracy theories." *Current directions in psychological science* 26(6):538–542.

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.

Fearon, James D. 1999. "Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance." *Democracy, accountability, and representation* pp. 55–97.

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.

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.

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

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

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

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

Harsch, Joseph. 1954. "State of the Nations Guatemalan Interlude." *Christain 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.

Immerman, Richard H. 1979. "Eisenhower and Dulles: Who Made the Decisions?" *Political Psychology* 1(2):21.

Izzo, Federica. 2022. "Ideology for the Future." *American Political Science Review* pp. 1–16.

Jeffreys-Jones, Rhodri. 2022. *Covert Action in the 1950s*. Oxford University PressOxford pp. 33–48.

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

King, Robert. 2017. *The Timor Gap, 1972-2017*. Vol. 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. 2021. *Meddling in the ballot box : the causes and effects of partisan electoral interventions*.

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.

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

McAuliffe, Mary S. 1981. "Eisenhower, the President." *The Journal of American History* 68(3):625.

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

Napier, Creed. 2023. *At Arm's Distance: How States Manage Escalation with Proxies and Denials*. Stanford University.

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.

Poznansky, Michael. 2019. "Feigning Compliance: Covert Action and International Law." *International Studies Quarterly* .

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.".

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

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.

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

Senter, William. 1963. "Memorandum from Major General W. O. Senter, Assistant Deputy Chief of Staff, Systems and Logistics, U.S. Air Force, to Distribution List, A Plan for the Withdrawal and Disposition of the SM-78 (Jupiter) Weapon System from Italy and Turkey:Operation Pot Pie.".

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 Nuclear Age Series.

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.

Times, New York. 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*

UN. 1954. "Cablegram dated 19 June 1954 from the Minister for External Relations of Guatemala addressed to the President of the Security Council. 14 June (S3232).". https://digitallibrary.un.org/record/602948.

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

7 Appendix

Outline:

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 otherwise.

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

$$Pr(z = 1 | a_c) = a_c \lambda, \qquad \lambda \in [0,1]$$

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

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

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

Altogether, this brings us to a formal articulation 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.

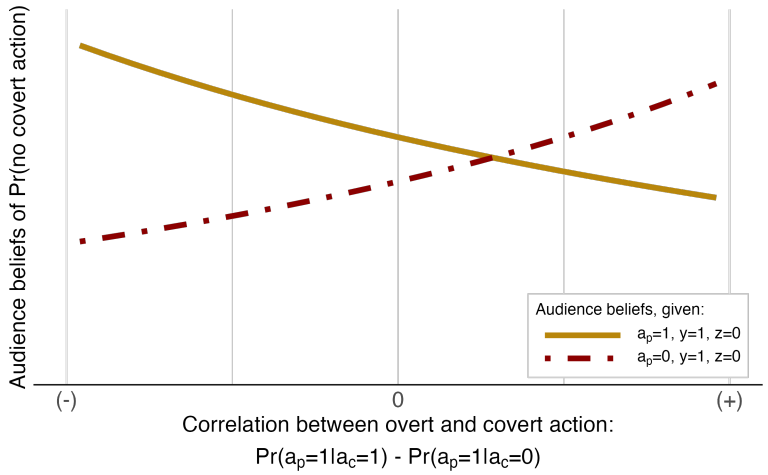
■

Figure 4 replicates Figure 1 from the main text (which featured addictive effects of covert and public action), along with two other cases featuring complements and substitutes. All three cases are generated with the paramater 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 first panel, with additive effects of public and covert action, we have $\alpha_0 = \alpha_p + \alpha_c - \alpha_{pc} = 0.25$. In the second panel, with substitutes, we have $\alpha_0 = 0.1$. In the third panel, with complements, we have $\alpha_0 = 0.4$.

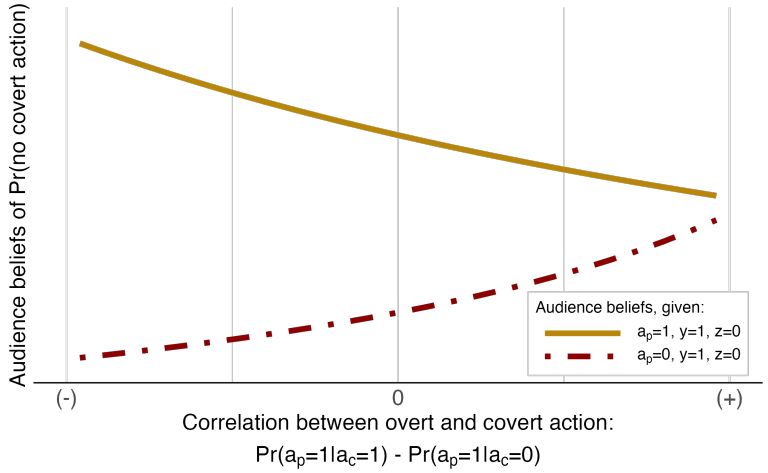
We label the horizontal axis as representing variation in the correlation between overt and covert action. To be more precise, the horizontal axis captures the difference in probabilities of public action conditional on covert action vs. covert inaction, $Pr(a_p = 1|a_c = 1) - Pr(a_p = 1|a_c = 0)$. This 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 $\sqrt{V[a_p]/V[a_c]}$.

Figure 4: Audience Beliefs in Non-Strategic Model: Comparing Across Interactive Effects of Covert and Public Action

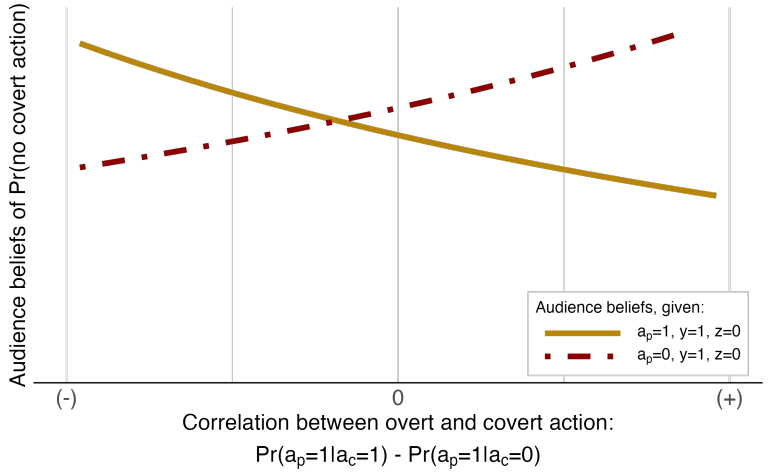
(a) Additive effects of covert and public action, $\alpha_c - \alpha_0 = \alpha_{pc} - \alpha_p$



(b) Covert and public action are substitutes, $\alpha_c - \alpha_0 > \alpha_{pc} - \alpha_p$



(c) Covert and public action are complements, $\alpha_c - \alpha_0 < \alpha_{pc} - \alpha_p$



7.2 Comments on Modeling Assumptions

Here we offer some 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 (particularly in the pure moral hazard setting, presented in Appendix 7.5). 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 in an intermediate range relative to the threshold, as per Assumption 1(iv)).

7.3 Recap of Formal Results

First, we state two additional technical assumptions that are imposed throughout the analysis:

Assumption 1 (Parameter restrictions)

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

(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 in an intermediate range:*

- *The lower bound of* π *is given by:*

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

- *The upper bound of* π *is defined implicitly, in the proof of Proposition 2.*

This assumption 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 (that is, an equilibrium in which the scrupulous leader does not "pander"; see discussion below). Point (ii) means that the probably of success due to exogenous factors is relatively small. When it is violated, the leader's plausible deniability problem becomes trivial, as the audience becomes too inclined to grant the leader the benefit of the doubt. 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. Finally, point (iv) implies that it is neither too easy nor too difficult for the leader to avoid punishment from the audience, meaning that the audience can meaningfully influence the leader's behavior.[22]

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

This assumption is not used to support any equilibrium of interest. Rather, it is only used to eliminate substantively unappealing equilibria which would depend on the audience imposing off-path punishment with no justification.

Next, we define a class of substantively appealing equilibria:[23]

[22] Specifically, the lower bound on π implies that A's belief in the RE with no CS satisfies $\mu^{110} \geq \frac{1}{2}$ (see Lemma 6). The upper bound implies that if the leader uses a cover story too frequently, then μ^{110} can fall below $\frac{1}{2}$.

[23] Note that Assumption 2 does very little work within an RE: all $z = 0$ information sets are on-path in an RE, so Assumption 2 reduces to assuming that any off-path belief following $z = 1$ assigns probability 1 to the leader being unscrupulous.

Definition 3 *A **responsive equilibrium (RE)** is an equilibrium in which the scrupulous leader takes public action if and only if the direct policy benefits outweigh the direct costs: that is, she plays $a_p = \omega$.*

Any behavior by the scrupulous leader which does not satisfy this condition would be a form of "pandering", a phenomenon detailed in previous work (Canes-Wrone, Herron and Shotts, 2001; Maskin and Tirole, 2004), but which is not the substantive focus of our analysis. Following the proof of Corollary 2 (in Appendix 7.4), we elaborate on the distinction between pandering and cover stories.

Proposition 2 (RE Existence) *A responsive equilibrium always exists.*

While the RE we characterize in the main text analysis are not necessarily unique, the following result provides a justification for focusing on the RE even when other equilibria can be supported.

Corollary 3 (RE Optimality) *Among all equilibria, the RE yields the best policy payoff for the scrupulous leader. If α_0 is low, the RE yields the best overall payoff for the scrupulous leader.*

We now restate the formal results from Section 3 of the main text (along with Proposition 3, which did not appear in the main text), before presenting the proofs for each.

Proposition 1 (Equilibrium Regions within RE). There exist thresholds λ^* and λ^{**} such that, within any RE:

- If $\lambda \geq \lambda^{**}$, the leader never takes covert action.
- If $\lambda^* < \lambda < \lambda^{**}$, the leader takes covert action with positive probability, but never uses a cover story.
- If $\lambda < \lambda^*$, the leader uses a cover story with positive probability; that is, any RE must be a CSE.

Proposition 3 (CSE Existence) *There exists a threshold λ^* such that a CSE exists if and only if $\lambda \leq \lambda^*$.*

Recall that Propositions 1 and 3 (and all other formal results) invoke the definition of a CSE from Definition 2, which focuses on a "pure" cover story: public action taken when the direct costs outweigh the direct benefits (i.e. when $\omega = 0$).

Note that Proposition 3 is a weaker but more general claim than Proposition 1: Proposition 3 is a statement of CSE existence (without restricting attention to RE); while Proposition 1 says that, for $\lambda < \lambda^*$, all RE are CSE.

Corollary 1 (CSE Comparative Statics). The threshold λ^* is increasing in α_c and in α_p^0, and decreasing in k_p; and if α_0 is small, it is increasing in β.

Corollary 2 (Cover Stories and Scrutiny). Consider any RE in which a cover story is played with positive probability (that is, when $\lambda < \lambda^*$). The audience's interim beliefs about the leader's scrupulousness upon observing public action (but before observing the outcome, or any revelation of covert action) are strictly less favorable than their interim beliefs upon observing no public action.

7.4 Proofs of Main Text Results, Section 3

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* $q = r^{010}\beta$, $s = r^{110}\beta$, $t = r^{100}\beta$, $v = r^{000}\beta$.
- *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* $F(x)$.

To preview the structure of the proofs:

- Lemmas 1 and 2 establish basic properties of the leader's and audience's best-responses.
- Lemmas 3 and 4 provide a general characterization of when cover stories can be used in equilibrium.
- Lemmas 5, 6, and 7 characterize the leader's and audience's strategies, and the audience's beliefs, within an RE.
- Proposition 2 builds on these lemmas to show that an RE always exists.
- Lemma 8 demonstrates that there exists a threshold λ^{**} such that, within any RE, the leader takes covert action with positive probability if and only if $\lambda < \lambda^{**}$.
- Lemma 9 demonstrates that if λ is below a threshold λ^*, then any RE must be a CSE.
- Proposition 1 follows directly from Lemma 8 and Lemma 9.
- Proposition 3 extends Lemma 9, to show that in *any* equilibrium, a CSE exists if and only if λ is below a threshold.
- Corollary 3 compares the scrupulous leader's expected payoff in the RE vs. any other equilibria that may exist.
- Corollary 1 characterizes how the λ threshold from Proposition 3 varies as a function of parameters α_c, k_p, and β.
- Corollary 2 characterizes the audience's interim beliefs after observing public action but before observing the policy outcome.

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

- L^1 *plays* $a_p = 1$

- *When* $\omega = 1$*:* L^0 *plays* $a = (1,1)$ *if* $k_c < \ddot{k}_c^1$*, and* $a = (1,0)$ *otherwise.*
- *When* $\omega = 0$*:* L^0 *plays* $a = (0,1)$ *if* $k_c < \tilde{k}_c^0$*, and* $a = (1,0)$ *otherwise.*

If $\hat{k}_p^0 < k_p < \hat{k}_p^1$ *and* $\tilde{k}_p^0 < k_p < \tilde{k}_p^1$*:*

- L^1 *plays* $a_p = \omega$
- *When* $\omega = 1$*:* L^0 *plays* $a = (1,1)$ *if* $k_c < \ddot{k}_c^1$*, and* $a = (1,0)$ *otherwise.*
- *When* $\omega = 0$*:* L^0 *plays* $a = (0,1)$ *if* $k_c < k_c^*$*, and* $a = (0,0)$ *otherwise.*

If $\hat{k}_p^0 < k_p < \hat{k}_p^1$ *and* $\tilde{k}_p^1 < k_p$*:*

- L^1 *plays* $a_p = 0$
- *When* $\omega = 1$*:* L^0 *plays* $a = (1,1)$ *if* $k_c < \hat{k}_c^1$*, and* $a = (0,0)$ *otherwise.*
- *When* $\omega = 0$*:* L^0 *plays* $a = (0,1)$ *if* $k_c < k_c^*$*, and* $a = (0,0)$ *otherwise.*

Proof of Lemma 1: All claims in the lemma follow directly from comparing L's expected payoff from each of her available actions.

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

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

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

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

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

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

■

Remark 4 *An RE requires* $\tilde{k}_p^0 \leq k_p \leq \tilde{k}_p^1$

To see why this is the case, recall that an RE is defined as an equilibrium in which L^1 plays $a_p = \omega$. If $k_p < \tilde{k}_p^0$, then L^1 plays $a = 1$ when $\omega = 0$. Conversely, if $k_p > \tilde{k}_p^1$, then L^1 plays $a = 0$ when $\omega = 1$.

Remark 5 *Comparing across thresholds defined in Lemma 1:*

- $\tilde{k}_p^\omega - k_p = E[U_L(a = (1,0))] - E[U_L(a = (0,0))] = \hat{k}_c^\omega - \ddot{k}_c^\omega = k_c^* - \tilde{k}_c^\omega$
- $\hat{k}_p^\omega - k_p = E[U_L(a = (1,1))] - E[U_L(a = (0,1))] = \hat{k}_c^\omega - k_c^* = \ddot{k}_c^\omega - \tilde{k}_c^\omega$
- $\tilde{k}_p^0 < \tilde{k}_p^1$
- $\hat{k}_p^0 < \hat{k}_p^1$

Lemma 2 *In every equilibrium:*

- *Given history* $h = (a_p, y, z)$*, the audience's best response satisfies*

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

- *Upon observing the direct revelation of covert action* ($z = 1$)*, 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, so the observation of $z = 1$ implies that $\mu = 0$ (either on-path by Bayes' Rule, or off-path given Assumption 2). ■

Lemma 3 L^0 *uses a cover story with positive probability only if* $k_p \leq \hat{k}_p^0$*.*

Proof: A cover story requires playing $a = (1,1)$. If $k_p > \hat{k}_p^0$, then $a = (0,1)$ strictly dominates $a = (1,1)$, as per Lemma 1. ■

Lemma 4 *In any equilibrium:*

- $\hat{k}_p^0 \leq k_p$*; so, a CSE requires* $k_p = \hat{k}_p^0$*.*
- *If* $k_p > \tilde{k}_p^1$*, then* $\hat{k}_p^0 < k_p$*.*

Proof of Lemma 4: For the first point: Suppose $k_p < \hat{k}_p^0$. That means $a = (1,1)$ strictly dominates $a = (0,1)$, and thus $a = (0,1)$ is never played on the equilibrium path. In any such equilibrium, L^1 is weakly more likely than L^0 to play $a_p = 0$, so the audience's posterior belief given $a_p = 0$ is at least π, meaning the audience plays $q = v = \beta$. But $\hat{k}_p^0(q = v = \beta) \leq \alpha_p^0(1-\alpha_c)$, which is less than k_p by Assumption 1 (iii)—contradicting the supposition that $k_p < \hat{k}_p^0$. Thus we have that in any equilibrium, $k_p \geq \hat{k}_p^0$; combining this claim with Lemma 3, we know that a CSE requires that $k_p = \hat{k}_p^0$.

For the second point: In any equilibrium with $k_p > \tilde{k}_p^1$, L^1 never plays $a_p = 1$ on the path of play, so we must have that $s = t = 0$.[24] So we have $\hat{k}_p^0(s = t = 0) \leq \alpha_p^0(1-\alpha_c)$, which again is less than k_p. ■

Lemma 5 (Leader strategies within an RE) *In any RE satisfying $\hat{k}_p^1 > k_p$, the following strategy profile is the leader's best-response to the audience's strategy (q, s, t, v):*

- *L^1 plays $a_p = \omega$ and $a_c = 0$.*
- *When $\omega = 1$, L^0 plays $a = (1,0)$ if $k_c > \ddot{k}_c^1$, and $a = (1,1)$ otherwise.*
- *When $\omega = 0$:*
  - *If $k_p > \hat{k}_p^0$: L^0 plays $a = (0,0)$ if $k_c > k_c^*$, and $a = (0,1)$ otherwise.*
  - *If $k_p = \hat{k}_p^0$: L^0 plays $a = (0,0)$ if $k_c > k_c^* = \hat{k}_c^0$, and otherwise mixes between $a = (1,1)$ and $a = (0,1)$ (playing $a = (1,1)$ with probability $\hat{\sigma}_p$ and $a = (0,1)$ with probability $1 - \hat{\sigma}_p$).*

Proof of Lemma 5: The scrupulous leader's strategy follows from the definition of RE. If that strategy is supported, then we know that $\tilde{k}_p^0 \leq k_p \leq \tilde{k}_p^1$. The unscrupulous leader's strategy then follows from Lemmas 1, 3, and 4. ■

Lemma 6 (Audience beliefs within an RE) *Consider the leader's RE strategy characterized in Lemma 5, where L^0 plays a cover story with probability $Pr(a_p = 1 | \omega = 0, a_c = 1, \theta = 0) = \hat{\sigma}_p \geq 0$. In this equilibrium, A's beliefs satisfy:*

$$\mu^{000} = \frac{\pi(1-\alpha_0)}{\pi(1-\alpha_0) + (1-\pi)[F(k_c^*)(1-\alpha_c)(1-\lambda) + (1-F(k_c^*))(1-\alpha_0)]} \geq \pi$$

$$\mu^{110} = \frac{\pi\tau\alpha_p^1}{\pi\tau\alpha_p^1 + (1-\pi)\left[\tau F(\ddot{k}_c^1)\alpha_{pc}^1(1-\lambda) + \tau(1-F(\ddot{k}_c^1))\alpha_p^1 + (1-\tau)F(k_c^*)(1-\lambda)\alpha_{pc}^0\hat{\sigma}_p\right]}$$

$$\mu^{100} = \frac{\pi\tau(1-\alpha_p^1)}{\pi\tau(1-\alpha_p^1) + (1-\pi)\left[\tau F(\ddot{k}_c^1)(1-\alpha_{pc}^1)(1-\lambda) + \tau(1-F(\ddot{k}_c^1))(1-\alpha_p^1) + (1-\tau)F(k_c^*)(1-\lambda)(1-\alpha_{pc}^0)\hat{\sigma}_p\right]}$$

If $\hat{\sigma}_p = 0$, then $\mu^{100} \geq \mu^{110} \geq \frac{1}{2}$.

[24] If L^0 also never played $a_p = 1$, so $a_p = 1$ is off-path and thus $s = t = \beta$ by Assumption 2, then we would have $\tilde{k}_p^1(s = t = \beta) = \alpha_p^1 - \alpha_0 + \beta - \alpha_0 q - (1-\alpha_0)v \geq \alpha_p^1 - \alpha_0 > k_p$, contradicting $k_p < \tilde{k}_p^1$.

Proof of Lemma 6: Generally, observe that

$$\mu^{a_p,y,z} = Pr(\theta = 1|a_p, y, z) = \frac{Pr(a_p, y, z|\theta = 1)\pi}{Pr(a_p, y, z|\theta = 1)\pi + Pr(a_p, y, z|\theta = 0)(1-\pi)}$$
$$Pr(a_p, y, z|\theta) = \sum_{\omega} Pr(a_p, y, z|\theta, \omega)Pr(\omega)$$
$$Pr(a_p, y, z|\theta, \omega) = \sum_{a_c} Pr(y, z|a_p, a_c, \theta, \omega)Pr(a_p, a_c|\theta, \omega)$$

The μ^h expressions in the lemma follow simply from applying these formulas, along with the strategy profile characterized in Lemma 5. To see that $\mu^{000} \geq \pi$, observe that $(1-\alpha_0) > (1-\alpha_c)(1-\lambda)$. To see that $\mu^{100} \geq \mu^{110}$ when $\hat{\sigma}_p = 0$, observe that

$$F(x)\frac{(1-\alpha^1_{pc})(1-\lambda)}{1-\alpha^1_p} + (1-F(x)) \leq F(x)\frac{\alpha^1_{pc}(1-\lambda)}{\alpha^1_p} + (1-F(x))$$

Finally, to see that $\mu^{110} \geq \frac{1}{2}$ when $\hat{\sigma}_p = 0$, observe the following:

- When $\hat{\sigma}_p = 0$, we know that $\mu^{100} > \pi$, because $F(x)(1-\alpha^1_{pc})(1-\lambda) + (1-F(x))(1-\alpha^1_p) < (1-\alpha^1_p)$. Thus in this equilibrium, A plays $t = r$.
- Rearranging the expression for μ^{110} when $\hat{\sigma}_p = 0$, we have that $\mu^{110} \geq \frac{1}{2}$ if and only if

$$\frac{\pi}{(1-\pi)} - 1 \geq F(\ddot{k}^1_c)\left[\frac{\alpha^1_{pc}}{\alpha^1_p}(1-\lambda) - 1\right] \tag{8}$$

- The lefthand side of this inequality is positive, given $\pi > \frac{1}{2}$. If the quantity in the square brackets is nonpositive, the inequality is satisfied. If that quantity is positive, that means that $\lambda < \frac{\alpha_c(1-\alpha^1_p)}{\alpha^1_{pc}}$.
- Turning to the expression for $\ddot{k}^1_c$, we can see that it is strictly increasing in s for any $\lambda < \frac{\alpha_c(1-\alpha^1_p)}{\alpha_{pc}}$. Thus if (8) is satisfied for $s = \beta$, then it is satisfied for any s.
- The lower bound for π given in Assumption 1 is equivalent to (8) with $s = t = \beta$ plugged into the expression for $\ddot{k}^1_c$.

Altogether, given the lower bound on π, it follows that for the strategy profile characterized in Lemma 5 with $\hat{\sigma}_p = 0$, A's belief satisfies $\mu^{110} \geq \frac{1}{2}$. ■

Lemma 7 *In any RE, $\hat{k}^1_p > k_p$.*

Proof of Lemma 7: Suppose that there exists an RE with $\hat{k}^1_p \leq k_p$. This RE differs from the RE characterized in Lemma 5 only in that the unscrupulous leader plays $a = (0,1)$ when $\omega = 1$ and $k_c < \tilde{k}^1_c$ (rather than $a = (1,1)$ or $a = (1,0)$, as in the previously characterized RE). This change has the effect of increasing μ^{110} and μ^{100}, which were both above $\frac{1}{2}$ in the Lemma 5 RE. Thus in this new RE we also have $s = t = \beta$. But $\hat{k}^1_p(s = t = \beta) \geq \alpha^1_p(1-\alpha_c) > k_p$, contradicting $\hat{k}^1_p \leq k_p$. Thus no RE with $\hat{k}^1_p \leq k_p$ can be supported. ■

Proof of Proposition 2 (RE Existence): Our strategy for proving Proposition 2 involves showing that the equilibrium characterized in Lemma 5 can always be supported. This requires showing that, given the specified strategy profile, the audience holds posterior beliefs which support a punishment/reward strategy (as per Lemma 2) that satisfies $\tilde{k}_p^0 \leq k_p \leq \tilde{k}_p^1$, and $\hat{k}_p^0 \leq k_p \leq \hat{k}_p^1$. If these conditions on A's strategy hold, then (per Lemma 1) we can see that the L strategy characterized in Lemma 5 is incentive-compatible.

We will prove the proposition by considering two cases: first, when a CS is not being played, and second, when a CS is being played.

Consider the strategy profile from Lemma 5 in which a cover story is not being played, meaning $\hat{\sigma}_p = 0$. Then from Lemma 6 we know that A's beliefs satisfy $\mu^{100} > \mu^{110} \geq \frac{1}{2}$, and $\mu^{000} > \frac{1}{2}$. Thus a strategy of $s = t = v = \beta$ is consistent with A's beliefs. Given the bounds on k_p provided in Assumption 1, we can see that $\tilde{k}_p^0(s = t = v = \beta) \leq k_p \leq \tilde{k}_p^1(s = t = v = \beta)$, and that $k_p \leq \hat{k}_p^1(s = t = v = \beta)$. The last incentive-compatibility condition needed to support an RE with no CS is that $k_p \geq \hat{k}_p^0(s = t = v = \beta)$; either this is satisfied, or RE existence can be shown to hold in the next case, with a CS played with positive probability.

Next, consider an RE with CS played with positive probability, meaning $\hat{k}_p^0 = k_p$. Here we derive the implicit expression for the upper bound on π that was introduced in Assumption 1. Recall that in the RE with $\hat{\sigma}_p = 0$, we have $\mu^{100} > \mu^{110} > \frac{1}{2}$. Also observe that both μ^{100} and μ^{110} are continuous and strictly decreasing in $\hat{\sigma}_p$, and continuous and strictly increasing in π. Thus if π is not too high (i.e. sufficiently close to μ^{110} when $\hat{\sigma}_p = 0$), then there exists a value of $\hat{\sigma}_p$ that satisfies $\mu^{100} > \mu^{110} = \frac{1}{2}$.

Let us suppose that L sets $\hat{\sigma}_p$ so that $\mu^{110} = \frac{1}{2}$. Then any s is a best response for A. An RE/CSE is supported if $\hat{k}_p^0 = k_p$ and $\tilde{k}_p^0 < k_p < \tilde{k}_p^1$. Given $v = \beta$, we have $\tilde{k}_p^0(v = \beta) \leq \alpha_p^0 - \alpha_0(1-\beta)$, which is less than k_p. To satisfy $\hat{k}_p^0 = k_p$, A can play

$$s = s^*(q) := \frac{k_p - \alpha_p^0(1-\alpha_c) + (1-\lambda)\beta\alpha_p^0(1-\alpha_c) + (1-\lambda)\alpha_c q}{(1-\lambda)\alpha_{pc}^0}$$

which we can see is positive.[25] Then to support the RE we need $k_p \leq \tilde{k}_p^1(t = v = \beta, s = s^*(q))$. To see that this is satisfied, observe:

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

This expression is $\geq k_p$, given the upper bounds on k_p and β stated in Assumption 1. ■

[25] If $s^* > \beta$, this rearranges to $\hat{k}_p^0(s = t = v = \beta) < k_p$, which means that the RE is supported with no CS, as per the previous case.

Lemma 8 (RE under high transparency) *There exists a threshold λ^{**} such that, within any RE, the leader takes covert action with positive probability if and only if $\lambda < \lambda^{**}$.*

Proof of Lemma 8: Consider an RE in which the leader never uses covert action. In this equilibrium, the audience holds beliefs $\mu^{a_p,y,z=0} = \pi$ for all a_p, y, and plays $q = s = t = v = \beta$ (as per Lemma 6). Since the leader never uses covert action, we know that

$$\underline{k_c} \geq k_c^*(q = v = \beta) = \alpha_c - \alpha_0 - \beta\lambda$$

which rearranges to

$$\lambda \geq \frac{\alpha_c - \alpha_0 - \underline{k_c}}{\beta} =: \lambda^{**}$$

Thus we have shown that in any RE, $Pr(a_c) = 0 \implies \lambda \geq \lambda^{**}$, and by contraposition, $\lambda < \lambda^{**} \implies Pr(a_c) > 0$.

Next we want to show that $\lambda \geq \lambda^{**} \implies Pr(a_c) = 0$. If $Pr(a_c) > 0$, it must be the case that $\underline{k_c} < k_c^*$. From Lemma 6, we know that $v = \beta$; and since k_c^* is increasing in q, we know that $k_c^* \leq \alpha_c - \alpha_0 - \beta\lambda$; so $\underline{k_c} < k_c^*$ implies $\lambda < \lambda^{**}$. Thus $Pr(a_c) > 0 \implies \lambda < \lambda^{**}$, and by contraposition, $\lambda \geq \lambda^{**} \implies Pr(a_c) = 0$. ■

Lemma 9 (CSE threshold within RE) *Within any RE:*

- *If $\lambda > \bar{\lambda}(q^*)$, the leader never uses a cover story, and*
- *If $\lambda < \bar{\lambda}(q^*)$, the leader uses a cover story with positive probability,*

where

$$\bar{\lambda}(q) := \begin{cases} 1 - \left(\frac{k_p - \alpha_p^0(1-\alpha_c)}{\alpha_c(\beta - q)}\right), & q < \beta \\ -999 & otw \end{cases}, \qquad \text{and} \qquad q^* = \begin{cases} 0, & \hat{q} < 0 \\ \beta, & \hat{q} > \beta \\ \hat{q} & otw \end{cases},$$

where $\hat{q}$ is the unique solution to

$$\mu^{010} = \frac{\pi\alpha_0}{\pi\alpha_0 + (1-\pi)\Big[F(k_c^*(q, v = \beta))\alpha_c(1-\lambda) + (1 - F(k_c^*(q, v = \beta)))\alpha_0\Big]} = \frac{1}{2}$$

Proof of Lemma 9: First, observe that the expression for μ^{010} denotes A's belief $Pr(\theta = 1 | a_p = 0, y = 1, z = 0)$ in the RE with no CS. It is strictly decreasing in q. If $\mu^{010}(q = 0) < \frac{1}{2}$, then A's best response is $q = 0$; if $\mu^{010}(q = \beta) > \frac{1}{2}$, then A's best response is $q = \beta$; otherwise, the equilibrium requires that $q = \hat{q}$, which is the unique q that solves $\mu^{010} = \frac{1}{2}$.

Suppose that we have an RE with no CS. This means $k_p \geq \hat{k}_p^0(s = t = v = \beta, q = q^*)$, which rearranges to $\lambda \geq \bar{\lambda}(q^*)$. Conversely, if $\lambda < \bar{\lambda}(q^*)$, then the RE with no CS cannot be supported, and any RE must be a CSE, with $k_p = \hat{k}_p^0$.

Alternatively, suppose we have an RE with CS played with positive probability. L^1's strategy does not change relative to the RE with no CS, whereas L^0's strategy shifts some probability from $a = (0, 1)$ to $a = (1, 1)$, which makes q weakly increase relative to the RE with no CS. In the CS,

we know $k_p = \hat{k}_p^0$, which rearranges to $\lambda = \bar{\lambda}(q')$ for some $q' \geq q^*$. Because $\bar{\lambda}$ is decreasing in q, we know that $\lambda = \bar{\lambda}(q')$ implies $\lambda \leq \bar{\lambda}(q^*)$. This gives us the contrapositive of the first bullet point in the lemma. ■

Proof of Proposition 3 (CSE Existence): Recall:

- CSE existence requires $k_p = \hat{k}_p^0$, which in turn implies $k_p \leq \tilde{k}_p^1$ (as per Lemma 4).
- Proposition 2 showed that the RE (with $\tilde{k}_p^0 < k_p < \tilde{k}_p^1$) always exists.
- Lemma 9 showed that the λ^* threshold in the RE is $\bar{\lambda}(q^*)$.

We will consider separate equilibrium cases of: $k_p < \tilde{k}_p^0$; $k_p = \tilde{k}_p^0$; and $k_p = \tilde{k}_p^1$. (The case of $\tilde{k}_p^0 < k_p < \tilde{k}_p^1$ is the RE, which again was examined in Lemma 9.) We will show:

- The λ^* threshold, below which a CSE is supported, is highest in the equilibrium with $k_p < \tilde{k}_p^0$, if such an equilibrium exists.
- If an equilibrium with $k_p < \tilde{k}_p^0$ does not exist, then the highest λ^* threshold (across all possible equilibria) is $\bar{\lambda}(q^*)$, as per Lemma 9,

First consider the case of $k_p < \tilde{k}_p^0$. With no CS, this equilibrium features $s = t = \beta$, and $q = v = 0$. To support the equilibrium with no CS, we require $k_p \geq \hat{k}_p^0$, which rearranges to

$$\lambda \geq \lambda' := 1 - \left(\frac{k_p - \alpha_p^0(1 - \alpha_c)}{\beta} \right)$$

Conversely, any equilibrium with $k_p < \tilde{k}_p^0$ must be a CSE if $\lambda < \bar{\lambda}'$. We can see that $\lambda' > \bar{\lambda}(q^*)$.

Next, consider an equilibrium with $k_p = \tilde{k}_p^0$, with no CS. Such an equilibrium must feature $q \leq \beta$ and/or $v \leq \beta$, and consequently, is only supported for $\lambda \geq \lambda''$, where $\lambda'' \leq \lambda'$. Also note that whenever an equilibrium with $k_p = \tilde{k}_p^0$ exists, an equilibrium with $k_p < \tilde{k}_p^0$ exists as well.

Finally, consider an equilibrium with $k_p = \tilde{k}_p^1$, with no CS. Compared to the RE with no CS, such an equilibrium must feature a weakly lower value of s and t, and weakly higher value of q and v—all of which make CS less appealing than in the RE, meaning CS will be supported for a narrower range of λ relative to the RE.

Altogether: if any equilibria with $k_p < \tilde{k}_p^0$ exist when $\lambda \leq \lambda'$, then we can say that a CSE exists if and only if $\lambda \leq \lambda'$; otherwise, a CSE exists if and only if $\lambda \leq \bar{\lambda}(q^*)$. ■

Proof of Corollary 1 (CSE Comparative Statics): Recall from Proposition 3 that λ^* is either equal to λ' or $\bar{\lambda}(q^*)$. In the case that $\lambda^* = \lambda'$, the result follows simply from taking partial derivatives. In the case that $\lambda^* = \bar{\lambda}(q^*)$, the comparative statics must account for both the direct effects of the parameters on $\bar{\lambda}$ (holding fixed q^*), and any indirect effects via q^*:

- k_p and α_p^0 only have direct effects.
- The direct and indirect effects of α_c work in the same direction.

- The direct and indirect effects of β work in opposite directions; but if α_0 is sufficiently small, then $q^* = 0$ for any $\beta > 0$, which shuts down the indirect effect.

■

Proof of Corollary 3 (RE Optimality): The claim that the RE is yields the optimal policy payoff for the scrupulous leader follows trivially from the definition of the RE (the equilibrium in which the scrupulous leader plays $a_p = \omega$). The general strategy for proving the rest of the corollary (regarding overall payoff for the scrupulous leader) will be as follows:

- Observe that in any equilibrium, by definition, L^1 optimizes her payoff subject to A's punishment/reward strategy $\sigma_A = (s, t, q, v)$.
- Consider two equilibria, featuring $\sigma'_A = (s', t', q', v')$, and $\sigma''_A = (s'', t'', q'', v'')$, with $s' \leq s''$, $t' \leq t''$, $q' \leq q''$, $v' \leq v''$, with at least one strict inequality, and either:
  - the information set corresponding the strict inequality is reached with positive probability in both equilibria; or
  - L^1's strategy differs across the two equilibria.
- Then, the equilibrium with σ''_A yields L^1 a strictly higher expected payoff than the equilibrium with σ'_A.

We will consider two cases of the RE: first, when $\lambda \geq \bar{\lambda}(q^*)$, in which case L^0 never uses a cover story;[26] and second, when $\lambda < \bar{\lambda}(q^*)$, in which case she uses a CS with positive probability.

In the RE with $\lambda \geq \bar{\lambda}(q^*)$, so a CS is never played: as established previously, A plays $s = t = v = \beta$ in this equilibrium. For another equilibrium to yield L^1 a higher payoff, it would have to be the case that the other equilibrium features a higher q. If α_0 is sufficiently low, then either (i) the unscrupulous leader never plays $a = (0, 1)$, so $q = \beta$, or (ii) the unscrupulous leader plays $a = (0, 1)$ with positive probability, so $q = 0$.

In the RE with $\lambda < \bar{\lambda}(q^*)$, so CS played with positive probability: as established previously, there exists an RE/CSE with $t = v = \beta$ and $s = s^*$. Compare this to the other possible equilibria that exist under the same parameter values (all of which have $q = 0$, given sufficiently low α_0):

- In an equilibrium with $k_p > \tilde{k}{}^1_p$, we have $s = t = 0$ (because the scrupulous leader never takes public action); so this is dominated by RE/CSE.
- In an equilibrium with $k_p \leq \tilde{k}{}^0_p$, we must have $v < \beta$ (because if $v = \beta$ then $\tilde{k}{}^0_p$ would be strictly less than k_p); in order to yield a higher payoff than the RE/CSE, we would need $s > s^*$; but given $v < \beta$, any s that would satisfy $k_p \geq \hat{k}^0_p$ would have to be lower than s^* (and any $s \geq s^*$ would yield $\hat{k}^0_p > k_p$, which we established above cannot hold in equilibrium).
- In the best-case equilibrium with $k_p = \tilde{k}{}^1_p$, we have $\hat{k}^0_p < k_p$,[27] and $v = \beta$, with L^1 mixing between $a = (1, 0)$ and $a = (0, 0)$ when $\omega = 1$. The condition that $k_p = \tilde{k}{}^1_p$ becomes

$$k_p - \alpha^1_p + \alpha_0 + (1 - \alpha_0)\beta = \alpha^1_p \tilde{s} + (1 - \alpha^1_p)\tilde{t}$$

[26] Whether or not we assume L^0 plays a cover story in the knife-edge case $\lambda \geq \bar{\lambda}(q^*)$ is irrelevant for this analysis, since her payoffs are equivalent in either case.

[27] Except for a knife-edge condition on the parameters, we cannot simultaneously have $\hat{k}^0_p = k_p = \tilde{k}{}^1_p$.

for some $\tilde{s}, \tilde{t}$. Comparing L^1's expected payoff in the RE/CSE against this alternative equilibrium: her payoffs are the same when $\omega = 0$; when $\omega = 1$, her payoff in the RE/CSE is $\alpha_p^1(1+s^*(q=0))+(1-\alpha_p^1)\beta-k_p$, and her payoff in the alternative equilibrium is $\alpha_0+(1-\alpha_0)\beta$. The conditions on β and k_p in Assumption 1 imply that the RE/CSE payoff is higher.

Comment: The condition that α_0 is low is a sufficient condition for RE optimality, not a necessary one. For substantive intuition, consider the following: Consider a pandering equilibrium in which L^1 never takes covert of public action, and in which the audience plays $s = t = 0$, $q = v = \beta$. Compare this pandering equilibrium to an RE with $s = t = v = \beta$, $q = 0$. L^1's expected utility from the pandering equilibrium is $\beta + \alpha_0$, while her expected utility from the RE is $\tau(\beta + \alpha_p^1) + (1-\tau)(\alpha_0 + (1-\alpha_0)\beta)$; the EU from the RE is higher if α_0 is small. Other sufficient conditions could be derived involving high τ or high α_p^1. ■

Proof of Corrollary 2 (Cover Stories and Scrutiny): The audience's interim beliefs upon observing $a_p = 1$, but before observing y or z, are given by

$$\mu^{interim;a_p=1} = Pr(\theta = 1|a_p = 1) = \frac{\pi\tau}{\pi\tau + (1-\pi)[\tau + (1-\tau)F(k_c^*)\hat{\sigma}_p]}$$

which is strictly less than π if a cover story is played with positive probability. ■

Comment: This finding helps to clarify the distinction between cover stories and "pandering" (Canes-Wrone, Herron and Shotts, 2001; Maskin and Tirole, 2004). In a pandering equilibrium, the leader selects a policy that she believes is in neither her nor the audience's policy interests (playing $a_p \neq \omega$, in our setting) because doing so is "popular"—that is, because the audience expects the scrupulous leader to behave that way, and thus rewards that behavior and punishes any other behavior. In our cover story equilibrium, in contrast, Corollary 2 says that taking the public action (as part of the cover story) is actually *unpopular*, relative to doing nothing at all. The leader relies on a cover story because it is a less-bad option than achieving her desired policy outcome through objectionable means with no alternative explanation.

Remark 6 *Within an RE/CSE,* $Pr(\omega = 1|a_p = 1, z = 0) \in (\tau, 1)$.

Proof of Remark 6:

$$Pr(\omega = 1|a_p = 1, z = 0) = \frac{Pr(a_p = 1, z = 0|\omega = 1)\tau}{Pr(a_p = 1, z = 0|\omega = 1)\tau + Pr(a_p = 1, z = 0|\omega = 0)(1-\tau)}$$
$$Pr(a_p = 1, z = 0|\omega = 1) = \pi + (1-\pi)(1 - F(\ddot{k}_c^1)\lambda)$$
$$Pr(a_p = 1, z = 0|\omega = 0) = (1-\pi)F(k_c^*)\hat{\sigma}(1-\lambda)$$

where $\hat{\sigma} = Pr(a_p = 1|\omega = 0, \theta = 0, a_c = 1)$. Algebraic rearrangement shows that $Pr(\omega = 1|a_p = 1, z = 0) \in (\tau, 1)$ for any $\hat{\sigma} > 0$, ■

Remark 7 *Within an RE/CSE, when the leader takes covert action in the* $\omega = 0$ *state, she uses a cover story with strictly interior probability:* $Pr(a_p = 1|a_c = 1, \omega = 0, \theta = 0) = \hat{\sigma} \in (0, 1)$.

Proof of Remark 7: Suppose $\hat{\sigma} = 1$. This means that the leader never takes covert action without public action. Thus the audience necessarily forms the belief that $\mu^{0,1,0} \geq \pi$, and their best response requires $q = \beta$. Given this audience strategy, the unscrupulous leader has an incentive to deviate to never playing a a cover story, $\hat{\sigma} = 0$.

That $\hat{\sigma} > 0$ follows from the definition of the CSE. ■

Intuitively, this tells us that the more likely a leader is to use a cover story, the less valuable it becomes to do so. This tension implies that any use of cover stories must be part of a completely mixed strategy, given the leader's strategic imperative to maintain unpredictability.

7.5 Pure Moral Hazard

In the main text, we focused on the distinction between scrupulous vs. unscrupulous leaders, and the audience's challenge of distinguishing between them (that is, a model of adverse selection). This follows a long tradition in the political agency literature, which holds that prospective concerns over screening agents by quality will generally dominate any incentive to impose punishments on a purely retrospective basis (Fearon, 1999; Ashworth, 2012). In this section, we show that our novel cover story mechanism still emerges in a model of "pure moral hazard", where all leaders are commonly known to be unscrupulous, and the audience is simply trying to design an incentive scheme that minimizes the leader's use of covert action.

The model setup is as follows:

- The game sequence follows Figure 2, with the exception that the leader is commonly known to be of type $\theta = 0$, i.e. all leaders are unscrupulous.
- Leader payoff is still given by (4).
- Audience payoff is: $U_A = -a_c$; that is, the audience simply seeks to minimize the leader's use of covert action.

Rather than A's strategy being pinned down by sequential rationality given her beliefs of L's type (as it was in the model presented in the main text), here we will instead look for equilibria that maximize A's payoff.

Notation. Let

$$k_c^\dagger = \min k_c^* = k_c^*(q = 0, v = \beta) = \alpha_c - \alpha_0 - \beta((1 - \alpha_0) - (1 - \alpha_c)(1 - \lambda))$$

and let

$$k_c' = \min \tilde{k}_c^0 = \tilde{k}_c^0(s = t = \beta, q = v = 0) = k_p - \alpha_p^0 + \alpha_c - \beta$$

We will slightly modify Assumption 1, replacing it with the following:

Assumption 3 (MH Parameter Restrictions)

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

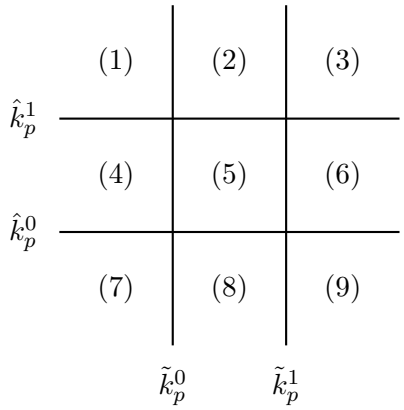
(ii) k_p *is in an intermediate range,* $\alpha_p^0 < k_p < \min\left\{\alpha_p^1 - \alpha_0, \alpha_p^1(1-\alpha_c)\right\}$

(iii) $\underline{k_c}$ *is in an intermediate range,* $\alpha_c(1-\alpha_p^1) - \beta\lambda \leq \underline{k_c} \leq k_c'$

The first two points are the same as in Assumption 1. Intuitively, the third point implies two things: (1) there exists an audience strategy which can fully disincentivize the leader from taking covert action when public action is effective ($\omega = 1$); and (2) there does not exist an audience strategy which can fully deter the leader from taking covert action by forcing her to always take public action.[28]

Lemma 1, Remark 5, and Lemma 3 remain unchanged. Thus we can partition the set of potential equilibria into nine cases, visualized in Figure 5. (Consider each boundary to be part of its interior case, e.g. case (4) denotes $\hat{k}_p^0 \leq k_p \leq \hat{k}_p^1$ and $k_p < \tilde{k}_p^0$.)

Figure 5: Equilibrium regions, as a function of k_p thresholds



Lemma 10 (CSE under pure moral hazard) *If* $\lambda < \bar{\lambda}(q=0)$*, as defined in Lemma 9, then there exists a CSE which yields* $Pr(a_c) \leq (1-\tau)F(k_c^\dagger)$.

Proof of Lemma 10: Propositions 2 and 3 demonstrated that when $\lambda < \bar{\lambda}(q=0)$, a CSE exists, characterized by: $t = v = \beta$, $q = 0$, and $s \in (0, \beta)$ that satisfies $k_p = \hat{k}_p^0 < \hat{k}_p^1$ and $\tilde{k}_p^0 \leq k_p \leq \tilde{k}_p^1$ (corresponding to case (5) in Figure 5). This was proven in the adverse selection setting, in which the audience's equilibrium strategy was pinned down by sequential rationality (i.e. maximizing (3)) given their beliefs about L's type. In the present setting, any audience strategy is permissible in equilibrium (including, of course, the audience strategy posited in Propositions 2 and 3). It is straightforward to see that the specified CSE is still supported with the additional restriction on $\underline{k_c}$ asserted in Assumption 3 (iii).

Plugging in the stated values of q, v, s, t, we can see that $k_c^* = k_c^\dagger$, and that $\ddot{k}_c^1 \leq \underline{k_c}$. This yields the equilibrium value of $Pr(a_c) = (1-\tau)F(k_c^\dagger) + \tau(0)$. ■

[28] Such an equilibrium would involve the audience fully rewarding whenever they observe public action ($s = t = \beta$), and fully punishing whenever they observe public inaction ($q = v = 0$), which (if $k_c' < \underline{k_c}$) would then ensure that the leader always takes public action.

Proposition 4 (Audience-optimal equilibria under pure moral hazard) *If $\lambda \leq \bar{\lambda}(q=0)$, then a CSE exists, and it achieves the highest possible audience payoff (i.e. the lowest possible $Pr(a_c)$) among all equilibria.*

Proof of Proposition 4: Following Lemma 10, all that remains to show is that when $\lambda \leq \bar{\lambda}(q=0)$, there does not exist any non-CSE equilibrium with $Pr(a_c) < (1-\tau)F(k_c^\dagger)$. We can consider each case depicted in Figure 5:

- Cases (7), (8), and (9) (all cases in which $k_p < \hat{k}_p^0$) are CSE.
- Case (1), with $\hat{k}_p^1 < k_p < \tilde{k}_p^0$, cannot be audience-optimal: within Case (1), $Pr(a_c)$ is strictly decreasing in s,t and strictly increasing in q,v, but $\hat{k}_p^1(s=t=\beta, q=v=0) > k_p$ when $\lambda < \bar{\lambda}(q=0)$.
- Case (2), with $\hat{k}_p^1 < k_p$ and $\tilde{k}_p^0 \leq k_p \leq \tilde{k}_p^1$, yields $Pr(a_c) \geq (1-\tau)F(k_c^\dagger)$.
- Case (3), with $k_p > \max\left\{\hat{k}_p^1, \tilde{k}_p^1\right\}$, yields $Pr(a_c) \geq F(k_c^\dagger)$.
- Case (6), with $k_p > \tilde{k}_p^1$ and $\hat{k}_p^0 \leq k_p \leq \hat{k}_p^1$, yields $Pr(a_c) \geq (1-\tau)F(k_c^\dagger)$.
- Among Case (5) equilibria, the lowest possible $Pr(a_c)$ is $(1-\tau)F(k_c^\dagger)$.
- Among Case (4) equilibria, with $k_p < \tilde{k}_p^0$ and $\hat{k}_p^0 \leq k_p \leq \hat{k}_p^1$: when $\lambda < \bar{\lambda}(q=0)$, any equilibrium with $k_p > \hat{k}_p^0$ cannot be audience-optimal. (Observe that in any Case (6) equilibrium with $k_p > \hat{k}_p^0$, $Pr(a_c)$ is strictly decreasing in s,t and strictly increasing in q,v; but $\hat{k}_p^0(s=t=\beta, q=v=0) > k_p$ when $\lambda \leq \bar{\lambda}(q=0)$.)

This exhausts all cases. ■