Cover Stories

Michael F. Joseph
UCSD
mfjoseph@ucsd.edu

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

January 22, 2025

Abstract

How do powerful states maintain plausible deniability for their secretive foreign interventions? Existing research focuses on the need for interveners to avoid *direct* exposure of their covert activities. We contribute by highlighting the challenges posed by *circumstantial* evidence: the inferences that audiences might draw from observing outcomes consistent with the intervener's interests and capabilities, but without direct evidence of their involvement. Through a series of formal models, we uncover a novel mechanism whereby interveners enhance their plausible deniability by openly promoting public coercive actions and, even, inviting some scrutiny that raises the risk of exposure. We find evidence of the use of such a "cover story" tactic in an in-depth case study of Operation PB-SUCCESS, the CIA operation that overthrew Guatemalan president Jacobo Arbenz in 1954. The theory advances our understanding of how leaders evade accountability for norm-violating behaviors across a wide range of foreign and domestic policy domains.

In March 1960, the CIA began organizing Cuban exiles to oust Fidel Castro from power. Eisenhower demanded that the CIA take extraordinary precautions to avoid direct evidence of U.S. involvement (Poznansky, 2020). But as CIA agents were secretly meeting Cuban contacts and building bases in Guatemala and Florida, Eisenhower initiated a public show-down with Castro. In December 1960, Eisenhower publicly announced a complete elimination of Cuba's sugar import quota, justified by Cuba's "deliberate hostility" towards the U.S. and increasing economic integration with the Soviet bloc (Eisenhower, 1960). The next month, the administration formally severed diplomatic ties with Cuba (DoS, 2023*a*)—a wholly symbolic gesture, as the U.S. ambassador had already been recalled and communication between the governments already ceased entirely (ADST, 2023, p.53-59). Shortly after, New York Times reports began speculating that the CIA could be training and equipping an invasion force.[1]

Why would Eisenhower invite scrutiny on an issue he wanted to keep secret? The conventional wisdom dictates that he would not. A substantial body of literature examines how governments maintain plausible deniability by avoiding *direct* evidence of their involvement in secret policies (Colaresi, 2012; Canfil, 2022; Carson, 2018; O'Rourke, 2018; Poznansky, 2020; Spaniel and Poznansky, 2018; Joseph and Poznansky, 2018; Carnegie, 2021). A common view across this literature is that overt action attracts scrutiny, which raises the risk that a secret policy is exposed.

The desire to avoid scrutiny is important, but it is not all that decision-makers 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 the audiences' ability to draw *strategic inferences* about 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 bring about that change. Rather, they also form inferences on the basis of *circumstantial evidence*, including their knowledge of the government's interests and capabilities, and of the broader context of the policy intervention. Even when the government succeeds in concealing direct evidence of its secret policies, it still faces a risk of audience backlash due to strategic inferences.

We argue that a government can maintain plausible deniability by using a *cover story*: an overt

[1] NY Times, Jan 6 and Jan 14, 1961

action that the governments can later rely on to explain how it 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 to the issue and raise the risk of exposure. But after the secret policy succeeds, those same public actions can reduce observers' retrospective suspicion over whether the policy success was achieved via secret means.

To navigate the complex set of tradeoffs and informational asymmetries inherent to this strategic setting, we analyze a series of formal models of an interaction between a government and an audience. In the models, the government can achieve its policy goals through different means—a public action, and a secret action—which can be taken in isolation or in combination. While the audience is broadly accepting of the government's policy objective, they find some means of achieving it more objectionable than others. The means that the audience finds most objectionable are the ones that the government pursues in secret. The audience's challenge is to prevent objectionable policies, given limited observability of the government's behavior. The government's challenge is to achieve its policy objectives, while avoiding reputational harm or reprobation from the audience—that is, while maintaining plausible deniability for any actions that the audience would disapprove of.

The analysis reveals that both taking public actions to generate a cover story, and avoiding public actions to avoid scrutiny, can (under different circumstances) constitute rational mechanisms to sustain plausible deniability. The key condition determining the government's optimal strategy is the baseline level of *transparency*: the probability that direct evidence of covert action will be revealed (Yoder and Spaniel, 2022). With sufficiently high transparency, the government is deterred from taking covert action entirely, as the risk of direct exposure outweighs the benefits of achieving the desired policy outcome. With moderate transparency, the government is willing to take covert action, in the hopes that direct evidence will not come to light, and the audience will attribute any successful outcome to exogenous factors outside of the government's control. When transparency is low, the government is still willing to pursue covert action, but it now finds that it must provide an alternative explanation for any successful outcome—by employing a cover story.

Why do the government's incentives change under conditions of low transparency? The reasoning follows from the audience's ability to draw strategic inferences of government behavior. If direct evidence of covert action is unlikely to come to light, then the absence of direct evidence

of covert action does not convince the audience that covert action was *not* taken. Rather, if they observe a successful outcome despite public inaction, they will simply infer that covert action was taken out of public view, and punish the government just the same as if direct evidence had been exposed. Under these conditions, the use of a cover story enables the government to achieve its desired policy ends while maintaining plausible deniability for the distasteful means.

From the government's perspective, the ideal cover story is an action which is taken publicly, and observed by the relevant audience; it plausibly contributes to achieving the policy objective being pursued through covert action; and it is viewed by the audience as less objectionable than the covert action being covered up. The cover story succeeds if it mitigates suspicion enough for the government to avoid backlash from the audience.

We examine this logic in an in-depth case study of Operation PBSUCCESS, Eisenhower's covert intervention to oust the Guatemalan President Jacobo Arbenz in 1954. It is well known that administration officials feared international backlash, and therefore only considered the mission successful if plausible deniability was achieved (Schmitz, 1999). Standard accounts show that the administration sought to avoid direct evidence of US involvement through tight operational controls, and by distancing themselves publicly from the coup plotters as the coup was ongoing.

Our analysis illuminates important events that are overlooked by existing accounts. We observe that the administration and CIA planners believed that many in Latin America would blame the US for Arbenz's removal even in the absence of direct evidence, because of strategic inferences. We argue that highly publicized actions by the U.S., including shipping embargoes, and sanctions and protests registered through the Organization of American States (OAS), cannot be fully explained by the administration's desire to use all available means to advance their objective. We further show that after the mission was complete, the US government highlighted their public actions as a cover story to disclaim responsibility for covert action—and that observers at the time found the cover story to be convincing. We supplement this in-depth case study with two shorter empirical vignettes, showing the theory's applicability to negotiations between Australia and East Timor over oil concessions in the Timor Gap, and to the resolution of the Cuban Missile Crisis.

Our theory enriches our understanding of a wide range of 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 overt actions can complicate attribution in ways that scholars have not previously considered. It also carries implications for research into international norms, laws and world order (Lake, Martin and Risse, 2021; Farrell and Newman, 2021). There is mounting evidence that states suffer costs for violating international laws and norms (Huth, Croco and Appel, 2011; Terman and Byun, 2022); but critics still worry that any constraining effects are undermined by the ability of powerful states to exploit covert action (Carson, 2018). Our theory suggests a practical limitation on how frequently states can exploit covert action to circumvent international responsibilities, especially if they do so repeatedly over time.

Our model also contributes to theoretical research on political agency and accountability (Ashworth, 2012). International relations scholars have shown that a principal-agent framework can be applied to a variety of settings in which one international actor seeks to influence another's behavior through the design of incentive schemes under incomplete information (Hawkins, Lake, Nielson and Tierney, 2006; Wolford and Rider, 2024; Ruachhaus, 2009; Biddle, Macdonald and Baker, 2018). We similarly demonstrate the framework's value for explaining how leaders can both be disciplined by the threat of punishments imposed by foreign or domestic audiences, and evade accountability for their secretive foreign policies. The broader political agency 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—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

Our analysis focuses on a setting that is commonly studied by scholars of secrecy and covert action (Spaniel and Poznansky, 2018; Poznansky, 2020; Carnegie, 2021). A government (the executive in particular) desires both a policy objective, and the approval of some relevant audience (e.g. domestic voters or legislators, foreign allies, or the broader international community). The

audience is generally accepting of that policy objective (e.g. preventing the spread of communism in the Western Hemisphere), but views some means to achieve it as more controversial than others (e.g. diplomatic and economic pressure vs. paramilitary operations and assassination attempts). If the government finds it infeasible to achieve the objective via the less controversial means, it may exploit secrecy to pursue the more controversial means, while attempting to conceal its action from the audience.

The incentive to exploit secrecy to achieve a policy success while avoiding political backlash arises in diverse policy contexts. For example, the U.S. public generally wants to control the flow of immigration at the southern border, but they do not want the government to achieve this goal by locking children in cages. The Trump administration initially sought to conceal its policy of family separation from the U.S. and foreign publics, and succeeded in keeping the policy secret for several months (Horowitz, 2021). When the policy was revealed it invited public condemnation, even from members of the Republican party (Todd, 2018). The U.S. public and European allies wanted President Kennedy to prevent Russia from deploying missiles to Cuba, but they did not want him to achieve this outcome by sacrificing the U.S. nuclear posture in Europe (Bernstein, 1980; Seneter, 1963). Thus, the administration concealed the missile exchange deal that facilitated the resolution of the Cuban Missile Crisis, even as they touted Soviet withdrawal as a success. The public broadly wants the government to make scientific advancements, but they do not want the government to achieve them through unethical experimentation. Thus, government scientists during the 1950s–1970s chose to administer unethical experiments on remote, marginalized communities—ethnic minorities, prisoners, and the mentally ill—hoping to reap the policy benefits while concealing unethical research practices that contributed to breakthroughs (ACHRE, 1996).

While the policy domain that plausibly fits our theoretical focus is broad, existing research on this sort of secrecy is primarily advanced by scholars of security and conflict studies, with a particular focus on covert intervention. Thus for concreteness, we will 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 controversial. The theoretical model is agnostic as to why the audience finds a policy objectionable; following previous studies of covert action, we often describe it as a policy that violates international laws and norms (Poznansky, 2019), or one that avoids institutional checks and balances from Congress (Smith, 2019) or procedural transparency to the

press and the public (Spaniel and Poznansky, 2018).

Consistent with existing research, we begin from the assumption that the success of any covert intervention depends critically on the ability of the Intervener to maintain *plausible deniability* for their actions. We depart in our assumptions about what plausible deniability requires. In existing theories, whether plausible deniability is maintained is treated as a deterministic function of the direct evidence that is revealed. This focus on direct evidence is far-reaching throughout the literature. 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.[2]

We argue that audiences are clever, and this creates a strategic barrier for sustaining plausible deniability that is not explored in existing research. 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 their expectations about how likely it is that those outcomes would occur in the absence of intervention.

The historical record is replete with important foreign policy decisions made on the basis of circumstantial evidence. For instance, in the late 1980s, Iranian dissidents living in Europe were much more likely murdered in a "robbery gone wrong" than the average European citizen. The German government indicted the IRCG absent any direct evidence of the IRCG's involvement (Hakakian, 2011). In 1950, the communist-leaning Bulgarian government foiled a coup plot. Given the U.S. position on communism, and U.S. Ambassador Donald Heath's broad personal relations throughout the Bulgarian political scene, the Bulgarian government inferred that Heath was involved in a covert operation. They declared Heath persona non-grata, leading U.S.-Bulgarian relations to sever (DoS, 2023*b*); with 70 years of hindsight, historians have not (to our knowledge) uncovered any evidence of U.S. involvement. Recent experimental work suggests that mass publics react similarly

[2]Note that our use of the term "circumstantial evidence" differs from Canfil (2022)'s. He uses the term to refer to *direct* evidence of actions that the government takes *indirectly*, i.e. through proxies, but does not treat the audience as a strategic actor. In contrast, our use refers to inferences formed by a strategic audience engaging in Bayesian updating.

to unproven and unclaimed coercive acts, demanding retaliation against the alleged perpetrator despite lacking direct evidence of their culpability (Pischedda and Cheon, 2023).

When the goal of plausible deniability is to avoid backlash for unscrupulous policies, mission success requires that Interveners convince relevant audiences, to a sufficient degree of confidence, that they were not responsible for the outcomes that result from those policies. As mission planning and execution is underway, Interveners must avoid direct evidence of their involvement. After the mission is complete, they must find a way to avoid strategic inferences of their culpability. The analysis that follows introduces the concept of a *cover story* as a tactic that Interveners can employ to maintain plausible deniability in the face of strategic inferences.

2 Formal Model

Our formal analysis focuses on a strategic interaction between a leader who seeks to advance a foreign policy objective through some combination of open and secret actions, and an audience who can punish or reward the leader for her behavior. This setting shares some core strategic similarities with principal-agent models, which study problems of designing incentives under incomplete information. While we develop our model specifically to study our substantive problem of interest, we do so using standard terminology from principal-agent theory in order to help relate our model to that broader literature and clarify its contributions relative to existing work.

We present two versions of the model, showing that our novel cover story mechanism arises in equilibrium under both. Each version reflects a different approach to modeling principal-agent relationships, and each highlights a different substantive aspect of the strategic problems inherent in sustaining plausible deniability over secret policies. The first model focuses on the leader's response to the incentive scheme imposed by the audience, considering whether the leader can evade accountability for authorizing covert action or whether she can be effectively deterred from it. Formally, this is a model of pure moral hazard with homogeneous leader types. It does not interrogate whether the audience is actually *willing* to impose the optimal incentive scheme, but rather focuses on the leader's behavior, taking the incentive scheme as given.

The second model highlights the audience's inferential challenge: how they form beliefs of the leader's (partially unobserved) actions, and what sort of punishments and rewards they are willing

to impose given those beliefs. Formally, this is a model of adverse selection with heterogeneous leader types, where a leader's "type" represents her intrinsic (un)willingness to abide by commonly shared international laws and norms, and the audience's payoff from "punishing" or "rewarding" the leader depends on her type. This model imposes slightly more restrictive assumptions than the first, and yields richer and more nuanced implications as a result.

After presenting the two models, we briefly consider two formal extensions motivated by the particular substantive setting of interest, showing the robustness of our proposed mechanism. Finally, we outline the key observable implications of the models, which we examine in the empirical analyses that follow.

2.1 Framework: Policymaking with public and secret options

To begin, we introduce the features of the model setup that are common across the two models. We study an interaction between a leader L of an Intervener state, and an audience A who can hold the leader accountable for her policy actions and outcomes. L can represent the leader acting alone, or in concert with her foreign policy advisory team (insofar as the latter is privy to the leader's choices). The audience can represent the Intervener's own electorate, or mass publics or political elites across different foreign countries—including the target of L's intervention, adversary states, or third-party states. While each of these audiences may have distinct relationships with the Intervener, and distinct preferences regarding the Intervener's behavior more broadly, our analysis focuses on a desideratum that we believe they all share: minimizing the Intervener's use of unscrupulous or norm-violating covert action.[3]

The sequence of moves is presented in Figure 1. The leader has two policy levers available, and she can enact either one, both, or neither. First is a public action $a_p \in \{0, 1\}$, which is taken openly. Second is a covert action $a_c \in \{0, 1\}$, which is taken secretly. We assume both actions are costly to implement (which could represent material, or moral costs), but that the covert action is more undesirable to the audience. For example, a_p may only exploit loopholes in international laws and weakly violate norms, whereas a_c could outright violate them. Referring back to our opening

[3]Recent experimental work by Myrick (2020) supports the claim that the U.S. public is generally opposed to covert foreign policy action by their own leader, all else equal. Notably, Myrick's study uncovers such a penalty for secrecy while holding the actions themselves fixed; in our analysis, the kinds of actions pursued covertly are inherently more objectionable than the kinds of actions pursued openly.

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

Figure 1: Game Sequence

anecdote for concreteness, a_p can represent the Eisenhower administration's imposition of economic pressure on Cuba through the slashing of sugar quotas and introduction of oil embargoes, while a_c can represent the various attempts made to oust or assassinate Castro through CIA-supported Cuban exiles or through agents operating secretly within the country.

The audience observes three pieces of information, which they can use to determine whether to reward or punish the leader. First, A observes whether L took the public action a_p. Second, A observes the policy outcome y. Third, A probabilistically observes the revelation of L's covert action a_c. We will consider each of these in turn.

We denote the leader's action $a = (a_p, a_c)$, where $a_i \in \{0,1\}$ for $i = p, c$ denote public action and covert action, respectively. The state variable $\omega \in \{0,1\}$ denotes whether the public action is *effective* or *ineffective*. If $\omega = 1$, the leader has reason to believe that the public action is likely to produce a successful outcome; if $\omega = 0$, the leader knows that the public action will not meaningfully advance the policy objective.

The policy outcome $y \in \{0,1\}$ (where $y = 1$ denotes success, $y = 0$ failure) is a probabilistic function the leader's action a, and the state variable $\omega \in \{0,1\}$. Formally, the policy technology

can then be represented as:

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

We assume that $\alpha_p > \alpha_c > \alpha_0$. Intuitively, this means that both covert action and (effective) public action can independently raise the probability of policy success, relative to the probability α_0 of a successful outcome arising due to exogenous factors (i.e., by random chance). The assumption that $\alpha_p > \alpha_c$ means that the leader has better (though still imperfect) *information* about the effectiveness of public action as compared to covert action.[4]

Note that this policy production function assumes complete technological substitutability between the two policy levers: if effective public action is being taken, then covert action has no additional impact on the outcome. This assumption serves to clarify how the concurrent use of public and covert action can arise as a result of *strategic* incentives, despite the technological disutility of doing so. An extension presented below demonstrates that our results are robust to relaxing this assumption, instead allowing for both public and covert action to contribute meaningfully to achieving the policy outcome.

Next, consider the covert action revelation technology. Specifically, let $z \in \{0,1\}$ denote whether covert action is revealed, with

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

Whenever the leader refrains from covert action, A observes $z=0$; but if the leader does take covert action, A observes $z=1$ with probability $\lambda \in (0,1)$. Thus λ denotes the risk of revelation of *direct* evidence of covert action; this is distinct from, though related to, the inferences audiences will draw on the basis of *circumstantial* evidence, as we will see in the analysis that follows.[5]

The last move of the game is A's decision to punish or reward the leader. A chooses an action $r \in [0,\bar{r}]$, with 0 denoting maximal punishment, and $\bar{r}$ denoting maximal reward; alternatively, we

[4]The total probability that public action will yield a successful outcome, $Pr(y=1|a_p\omega=1) \times Pr(\omega=1) = \alpha_p\tau_1$, may still be less than the corresponding probability for covert action, $Pr(y=1|a_c=1) = \alpha_c$.

[5]In an extension we assume taking public action raises the risk of direct exposure by making the total exposure risk equal to $a_c(\lambda + a_p\delta)$. We still find a cover story.

can understand $\bar{r}$ as denoting L's sensitivity to whatever maximal punishment or reward A is able to impose. Because A only observes $(a_p, y, z) \in \{0,1\}^3$, they can only condition their strategy on these observable pieces of information.

Finally, recall that the leader privately observes two cost variables, k_p and k_c. The variable $k_c \sim U[\underline{k_c}, \overline{k_c}]$ denotes the direct cost of taking covert action, and $k_p \in \{\underline{k_p}, \overline{k_p}\}$ denotes the direct cost of taking public action. If the public action is effective ($\omega = 1$) but prohibitively costly ($k_p = \overline{k_p}$), the leader will not pursue it.[6] If the public action is ineffective ($\omega = 0$) but not prohibitively costly ($k_p = \underline{k_p}$), the leader may still find it strategically optimal to take the action. Importantly, only the leader observes the realization of these two cost variables (and the state variable), while the audience maintains a prior belief that[7]

$$Pr(\omega = 0, k_p = \underline{k_p}) = \tau_0, \quad Pr(\omega = 1, k_p = \underline{k_p}) = \tau_1, \quad Pr(k_p = \overline{k_p}) = 1 - \tau_0 - \tau_1$$

We will note that the possibility of $k_p = \overline{k_p}$, occurring with probability $(1 - \tau_0 - \tau_1)$, is primarily introduced for technical convenience, to ensure that certain information sets remain on the equilibrium path of play. This probability can be arbitrarily small, and is not a substantive focus of the analysis.

Altogether, the leader's payoff is given by

$$U_L = y - a_c k_c - a_p k_p + r$$

The audience's payoff will vary across the two models, as we discuss below.

Throughout the analysis, we maintain the following parameter restrictions:

- $\bar{r} < 1$: the leader values the policy outcome more than whatever reward/punishment the audience can impose on her.
- $\alpha_0 < \min\{\underline{k_p}/\bar{r},\ \alpha_c(1-\lambda)\}$: the likelihood of policy success due to exogenous factors is relatively small.

[6] Formally, we assume $\overline{k_p} > (\alpha_p - \alpha_0) + \bar{r}$.

[7] Note that when $k_p = \overline{k_p}$, it is irrelevant whether $\omega = 0$ or $\omega = 1$, as the public action is never taken; thus we can simply state that $Pr(k_p = \overline{k_p}) = Pr(k_p = \overline{k_p}, \omega = 1) + Pr(k_p = \overline{k_p}, \omega = 0) = 1 - \tau_0 - \tau_1$ without differentiating between the two conditions.

- $\underline{k_p} \leq k_p^{\dagger} = \alpha_c(\alpha_p - \alpha_c + \underline{k_c} + \bar{r})$: this ensures that the leader is willing to take public action in the favorable state, when she is not punished for doing so.

2.2 The Cover Story

The core mechanism of interest in this analysis is the leader's use of a *cover story*, which we formally define as follows:

Definition 1 *A leader employs a* ***cover story*** *when she takes an ineffective public action ($a_p = 1$ when $\omega = 0$) while also taking covert action ($a_c = 1$).*

Our framework is deliberately set up so that if $\omega = 0$, public actions have no direct value in increasing the probability of policy success, and impose some direct cost on the leader. While this assumption is somewhat extreme, it provides the clearest exposition of our proposed mechanism, in that it cleanly distinguishes the direct policy value of public action from the indirect effect of public action on audience beliefs.

To make this definition concrete, let us return to the example of Eisenhower's Castro policy. Now-declassified documents indicate that the administration privately knew that the Soviets were willing and able to mitigate the impacts of both US oil embargoes[8] and sugar quotas. As such, NIE 85–2–60 argued that "Fidel Castro will almost certainly remain in power through 1960", despite the overt policies being pursued (CIA, 1960). Important for our argument is the notion that these intelligence assessments were *private*, and that some relevant foreign and domestic audiences faced uncertainty as to whether these sorts of policies could in fact contribute to the downfall of the Castro regime. In this context, we would say that the economic policy actions served as a cover story, which was employed alongside the more objectionable covert actions which the administration sought to keep hidden.

In the analysis that follows, we show that a cover story arises as equilibrium behavior, under the same conditions, across two different models of the interaction between the leader and audience.

[8]See: Memorandum of Discussion at the 450th Meeting of the National Security Council, Washington, July 7, 1960

2.3 Model 1: Leader's response to the audience's optimal incentive scheme (pure moral hazard)

The first model analyzes the moral hazard problem between the leader and the audience. The audience's goal is is to develop a scheme of punishments and rewards that elicits their most preferred behavior from the leader—namely, minimizing the leader's use of covert action. Importantly, while the behavior that the audience wants to disincentivize is (with some probability) unobservable, she can only condition her punishments and rewards on *observable* behaviors and outcomes. As we will see, even the optimal punishment scheme for the audience thus creates an opportunity for the intervener to exploit her informational advantage, by using a cover story.

2.3.1 Technical Assumptions

The audience's payoff is given by

$$U_A = -a_c \tag{3}$$

This assumes that the audience simply seeks to minimize the leader's use of the objectionable policy. While other considerations could be incorporated into the audience's payoff to increase the model's descriptive realism—for instance, assuming that the audience also places some value (positive or negative) on the foreign policy outcome itself—such modifications would distract from and obscure the core causal mechanism that our analysis aims to elucidate (Paine and Tyson, 2020).

Because the audience moves last, and their payoff is fully determined by the time that they makes their move, any punishment/reward scheme is sequentially rational and can be part of an equilibrium (as long as the leader is best-responding to whatever the audience's strategy is). This is a standard feature of pure moral hazard models.[9] Such models generate predictions through an equilibrium selection rule that restricts attention to the equilibrium that is optimal from the principal's (audience's) perspective.

In our setting, this standard selection rule is insufficient to yield determinate predictions, so we impose an additional selection rule of Pareto optimality. Specifically:

Assumption 1 (Equilibrium selection) *Restrict attention to the Subgame Perfect Equilibria that yield the highest payoff for the audience. Among these equilibria, restrict attention to the*

[9]See Barro (1973); Ferejohn (1986); Bueno de Mesquita (2007); Ashworth and Ramsay (2024)

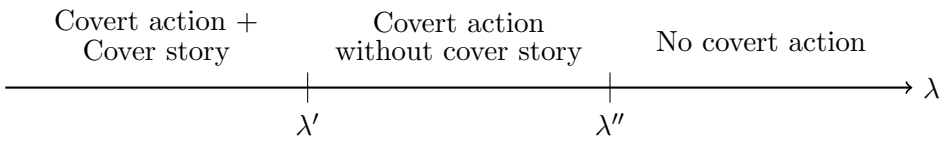


Figure 2: Equilibrium regions, as characterized in Proposition 1

equilibria that yield the highest payoff for the leader.

2.3.2 Analysis

Since A's objective is to prevent covert action, their equilibrium strategy must impose the maximum punishment if they observe direct evidence of covert action. It naturally follows that the underlying risk of direct exposure, represented by the parameter λ, will be fundamental in determining L's equilibrium strategy. Thus we can characterize the model's equilibrium as a function of cutpoints of this parameter.

Proposition 1 (Equilibria under pure moral hazard) *In the pure moral hazard game, there exist thresholds λ' and λ'' of the transparency parameter λ such that:*[10]

- *If $\lambda \geq \lambda''$, then in any equilibrium satisfying Assumption 1, the leader never takes covert action.*
- *If $\lambda \in (\lambda', \lambda'')$, then in any equilibrium satisfying Assumption 1, the leader takes covert action with positive probability but never uses a cover story.*
- *If $\lambda < \min\{\lambda', \lambda''\}$, there exists an equilibrium satisfying Assumption 1 in which the leader uses a cover story with positive probability.*

A simple depiction of equilibrium behavior is presented in Figure 2.

Let us consider the equilibrium logic within each of these regions. It is straightforward to see that any equilibrium that maximizes the audience's payoff will feature the audience fully punishing the leader upon observing direct evidence of covert action ($z = 1$), since this is exactly the behavior that the audience seeks to disincentivize. When transparency is sufficiently high ($\lambda \geq \lambda''$), this

[10] As derived in the appendix, $\lambda' = 1 - \underline{k_p}/(\alpha_c \bar{r})$ and $\lambda'' = \frac{(\alpha_c - \alpha_0)(1-\bar{r}) - \underline{k_c}}{\bar{r}(1-\alpha_c)}$.

risk of direct exposure, combined with the audience's maximal punishment in the event of direct exposure, is sufficient to completely deter the leader from taking covert action.

When transparency decreases below λ'', the leader finds the risk of covert action to be (sometimes) worthwhile (with randomness arising from the leader's privately-known direct cost of covert action, k_c). In this case, while the audience cannot deter covert action completely, they can still discourage it on the margin. The audience's optimal strategy for doing so punishes the leader when they observe direct evidence of covert action ($z = 1$); punishing when they observe policy success without public action ($a_p = 0, y = 1, z = 0$); and rewarding when they observe neither public action nor policy success ($a_p = 0, y = 0, z = 0$). This strategy, while optimal among the audience's feasible strategies, will nonetheless involve punishing both too often (when successful outcomes occur by random luck, with probability α_0) and not often enough (when covert action is unsuccessful but remains unexposed). Such inefficiency is a standard feature of principal-agent models, in which the principal can only condition rewards and punishments on observable actions and outcomes.

If the audience's goal is to deter covert action, and they are indifferent as to whether or not the policy succeeds, then why not maximally punish the leader whenever they observe a policy success—or, for that matter, any public action at all? Specifically, consider an audience strategy of: reward if the leader does nothing and the policy fails ($a_p = 0, y = 0, z = 0$) and punish otherwise. If the leader were primarily concerned with her reputation, such an audience strategy could effectively discipline the leader into complete inaction. However, if the leader places sufficiently high value on the policy outcome, she will risk reputational harm in order to achieve policy success, taking whichever action carries the most favorable balance of benefits and costs. Under the proposed audience strategy, public action is punished unconditionally, while covert action is only punished conditional on policy success or direct exposure. This strategy thus has the inadvertent effect of incentivizing covert action over public action, even when public action is effective (i.e. when $\omega = 1$).

Therefore, the audience's optimal strategy under moderate transparency rewards the leader for taking public action, regardless of the outcome. This is not because the audience wants to incentivize public action per se, but rather because they find it a less-bad option than covert action. Such an incentive scheme is necessary to induce the leader to forgo covert action in favor of public action when both levers are available and effective.

The audience's strategy of rewarding public action while punishing successful covert action

provides the basis for the leader's incentive to use a cover story—taking ineffective public action (in the $\omega = 1$ state) while also taking covert action, in the hope that the covert action will remain unexposed and the audience will attribute the policy success to the observed public action. In order for the leader to use a cover story, however, the direct cost of public action, $\underline{k_p}$, cannot exceed the expected improvement in reward from the audience. This expected improvement is represented by $\bar{r}$ in the event that covert action is successful and unexposed, which occurs with probability $\alpha_c(1-\lambda)$. Observe that $\underline{k_p} > \bar{r}\alpha_c(1-\lambda)$ rearranges to $\lambda > \lambda'$, which is the "low transparency" condition in the proposition.

If λ goes below λ', and the audience's strategy remains the same, then the gambit of the cover story proves net beneficial for the leader. Understanding this, the audience must then reduce the reward they give following public action, down to the point that the leader is held indifferent between using a cover story and using covert action alone. Thus we observe another source of inefficiency, whereby the audience imposes a partial penalty on public action, regardless of whether it is actually accompanied by covert action or not.

As a technical point, note that Proposition 1 allows for cover stories to be played with positive probability, but does not require it. The following corollary, however, suggests why we should find it unlikely that cover stories will *never* never played in equilibrium, under conditions of low transparency:

Corollary 1 *If $\lambda < \min\{\lambda', \lambda''\}$, there exists a continuum of equilibria in which L plays a cover story with any probability $p \in [0,1]$. Suppose the players select the equilibrium according to some continuous distribution over equilibria. Then there is zero probability of selecting an equilibrium in which the leader never uses a cover story.*

A separate intuition for why we should expect cover stories to be used with strictly positive probability emerges from the alternative model setup with heterogeneous leader types, as discussed next.

2.4 Model 2: The audience's inferential challenge (adverse selection)

The pure moral hazard model provides a useful analytical benchmark, in that it demonstrates how a cover story can emerge in equilibrium, even under the audience's most favorable conditions.

However, pure moral hazard models rests on a strong assumption of perfect homogeneity among agents (Fearon, 1999). If agents differ in any way that is relevant to the principal's payoff, "the procedure used to solve pure moral hazard models falls apart" (Ashworth, 2012, 186). The principal's desire to punish or reward agents for past behavior is fully overwhelmed by the need to prospectively screen agents by quality.

There is good substantive reason to consider heterogeneity across leaders in this context. Scholars have long argued that powerful states take covert action to avoid a reputation as norm violators, or as hypocritically imposing institutional rules on others that they refuse to follow themselves (Bull, 2002; Poznansky, 2019; O'Rourke, 2018; Carnegie, 2021). The possibility of forming such reputations translates to the technical assumption that leaders come in different "types", with different sensitivities to such violations. How can leaders avoid developing negative reputations from their use of controversial covert actions?

We address this question by incorporating the strategic challenge of adverse selection into our framework. In this model, the leader of the intervening state may vary along a dimension that we refer to as *scrupulousness*: this quality captures the extent to which a leader has internalized the international norms and institutional commitments that her state purports to uphold, or her willingness to violate them when it is convenient to do so. Holding fixed the nature of the foreign policy issue, the effectiveness of different policy options, and the risk of audience backlash, we say that scrupulous and unscrupulous leaders differ in the intrinsic value they place on adhering to international laws and norms.

The audience's payoff from "punishing" or "rewarding" the leader depends on the leader's type. Substantively, the audience's action can be interpreted as choosing whether to cooperate with the government on future foreign policy initiatives (in the case of a foreign audience), or choosing whether to support the leader politically (in the case of a domestic audience). Thus rather than designing a punishment scheme that elicits optimal behavior from the Intervener, the audience is constrained by the need to make prospectively optimal decisions given the limited information they have available. We find that even when imposing this constraint on the audience's strategy, we uncover equilibrium behavior which is qualitatively similar to that of the pure moral hazard model: specifically, we find that the use of cover stories arises in equilibrium under low levels of transparency.

2.4.1 Technical Assumptions

The key technical difference between this model and the previous one pertain to the payoff functions of the leader and the audience. All other features of the model setup are the same as in the previous model.

The leader is either "scrupulous" or "unscrupulous", and we model the scrupulous leader as a behavioral type:[11] we assume that she never takes covert action, and takes public action if and only if the state is favorable (that is, $\omega = 1$ and $k_p = \underline{k_p}$). This simplifies the analysis, focusing our attention on how the audience's goal of selecting "good" types can effectively discipline the behavior of "bad" types, or fail to do so.

The unscrupulous leader's payoff is as stated in the previous model. The audience's payoff is given by

$$U_A = r\theta + (1 - r)\bar{\mu} \tag{4}$$

where $\theta = 1$ denotes that the leader is scrupulous, and $\theta = 0$ denotes unscrupulousness. This variable is realized at the start of the game and observed privately by L, with the audience holding a prior belief that $Pr(\theta = 1) = \pi$. The parameter $\bar{\mu}$ represents the "reservation payoff" that A receives from foregoing future cooperation with L (as an international audience), or from withdrawing political support from L (as a domestic audience); more directly, it determines the level of suspicion that A would have to hold regarding L's unscrupulousness before being willing to impose punishment. We assume that $\bar{\mu} < \pi$, 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.

2.4.2 Analysis

Under this alternative model setup, we obtain the following results:

Proposition 2 (Equilibria under adverse selection) *In the adverse selection game, there exist thresholds λ' and $\bar{\lambda}$ of the transparency parameter λ such that:*[12]

- *If $\lambda \geq \bar{\lambda}$, there exists an equilibrium with zero probability of covert action.*

[11] See Ashworth and Bueno de Mesquita (2014) for a similar approach.

[12] As derived in the appendix, $\lambda' = 1 - \underline{k_p}/(\alpha_c \bar{r})$, and $\bar{\lambda} = (\alpha_c - \alpha_0 - \underline{k_c})/\bar{r}$.

- *If $\lambda' < \lambda < \bar{\lambda}$, then in any equilibrium, the leader takes covert action with positive probability, but never uses a cover story.*
- *If $\lambda < \min\{\lambda', \lambda''\}$, and if $(\pi - \bar{\mu})$ is not too large, then:*
  - *There exists an equilibrium in which L plays a cover story with positive probability, and this equilibrium yields the lowest possible probability of covert action.*
  - *Any equilibrium featuring the lowest possible probability of covert action requires that a cover story be played with strictly positive probability.*[13]

We can gain a deeper understanding of the equilibrium logic by examining the audience's beliefs, and how those beliefs vary as a function of the transparency parameter λ. First consider the audience's belief upon observing policy success and public inaction, but no direct evidence of covert action: that is, in the information set $(a_p = 0, y = 1, z = 0)$, visualized by the solid red line in Figure 3. In this information set, the audience infers that one of two things must have happened: either the leader took successful covert action which went unexposed, or the policy success arose due to exogenous factors, without any intervention from the leader.

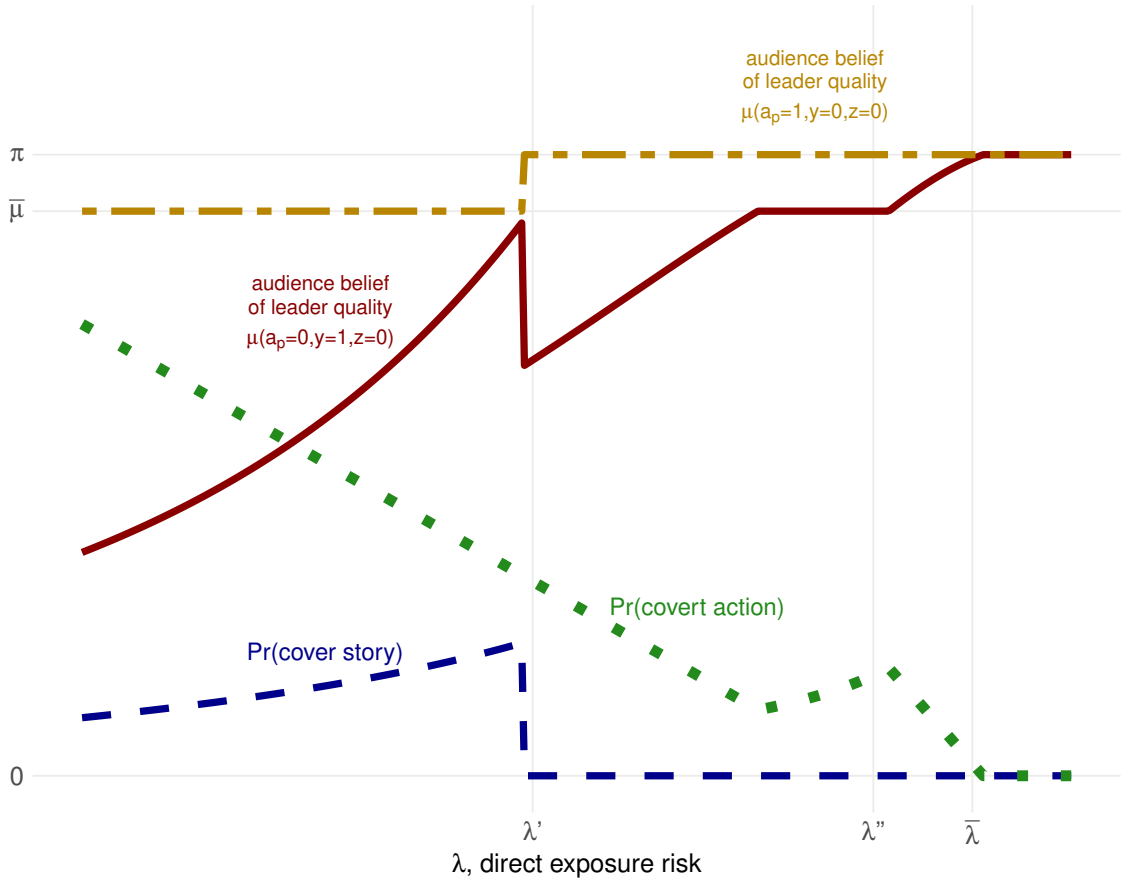
Under high transparency ($\lambda > \lambda''$), the direct risk of exposure causes the leader to take covert action only rarely (the dotted green line in the figure). Knowing this—and knowing that had covert action been taken, it likely *would have* been exposed—the audience is inclined to give the leader the benefit of the doubt in the $(a_p = 0, y = 1, z = 0)$ information set. This is reflected in the red line remaining above the punishment threshold $\bar{\mu}$. As transparency decreases, so too does the audience's belief of the leader's scrupulousness in this information set: 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 places more weight on covert action rather than random luck being the cause of the observed policy success. Eventually their belief drops below $\bar{\mu}$, and they fully punish the leader purely on the basis of circumstantial evidence.

Next, consider the audience's belief of the leader's scrupulousness upon observing public action, depicted by the dot-dashed gold line.[14] Across the range of $\lambda > \lambda'$, the audience's belief remains

[13] Further, note that in such an equilibrium, if $(1 - \tau_0 - \tau_1)$ is not too large, then L must play a cover story as part of a completely mixed strategy. That is: conditional on taking covert action in the $(\omega = 1, k_p = \underline{k_p})$ state, L must place positive probability on both $a = (0, 1)$ and $a = (1, 1)$.

[14] The belief given public action and policy failure, $\mu(a_p = 1, y = 0, z = 0)$, is pictured; the belief given public

Figure 3: Leader Strategy and Audience Beliefs



Note: Figure represents leader strategies and audience beliefs in an equilibrium satisfying the conditions of Proposition 2. Notation: π is the prior probability the leader is scrupulous; $\mu(a_p, y, z)$ denotes the audience's posterior belief; $\bar{\mu}$ is the audience's threshold for punishing the leader (see (4)). Figure is constructed with the following parameter values: $\alpha_p = 0.55, \alpha_c = 0.5, \alpha_0 = 0.04, \underline{k_c} = 0.05, \overline{k_c} = 0.5, \bar{r} = 0.5, \pi = 0.55, \bar{\mu} = 0.5, \underline{k_p} = 0.15, \tau_1 = 0.3$.

above $\bar{\mu}$. This is intuitive: given that the leader does not employ cover stories, and thus does not use public and covert action simultaneously, the audience has no reason to suspect covert action when they have observed that public action was taken. But it is precisely this favorable inference which the leader is then tempted to exploit, by using a cover story. As before, this requires that the direct cost of ineffective public action not exceed the reputational benefits, which is the condition that $\lambda \leq \lambda'$.

Here we observe a notable distinction between the equilibrium behavior of this model, and that of the pure moral hazard model analyzed previously: in this version, not only are cover stories

action and policy success, $\mu(1, 1, 0)$, remains above $\mu(1, 0, 0)$ for all λ values in this equilibrium, and is omitted for visual clarity.

possible in the audience-optimal equilibrium, but in fact the equilibrium *requires* that cover stories be used with positive probability. To see why, suppose that $\lambda < \lambda'$ and the leader never uses cover stories. Then sequential rationality would require that the audience maximally reward the leader whenever they observe public action, given their favorable belief in this information set. This would create an incentive for the leader to *always* use cover stories, which in turn would cause the audience to punish public action and reward public inaction. A stable equilibrium requires that the leader sometimes uses cover stories, and the audience partially punishes her for doing so.

A substantive implication of this analysis is that there exists fundamental limitation on the leader's ability to skirt accountability through cover stories. Inevitably, the use of cover stories is partially self-defeating: when the audience expects the leader to use cover stories, they become increasingly skeptical of successful outcomes that are accompanied by public action, ultimately making covert action alone a more attractive option for reputational reasons. Put differently, 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. As a more technical point, this result highlights the distinction between the homogeneous-types (pure moral hazard) model and the heterogeneous-types (adverse selection) model: by imposing a more stringent requirement on the audience's equilibrium behavior, this setup provides a more determinate prediction of the leader's behavior in response.

It is worth pausing to clarify exactly how the leader benefits from using a cover story. Perhaps surprisingly, the audience's interim beliefs—after observing the leader's action but before observing the outcome—are actually *less* favorable given public action, as compared to public inaction:

Corollary 2 (Cover Stories and Scrutiny) *For $\lambda < \lambda'$: Consider the equilibrium characterized in Proposition 2, featuring the lowest possible probability of covert action, and a cover story played with positive probability. After observing public action ($a_p = 1$), but before observing the policy outcome (y) or the covert revelation (z):*

- *The audience increases their belief that the leader is unscrupulous (as compared their belief after observing no public action).*
- *If the equilibrium probability of a cover story, $Pr(a = (1,1) \mid \omega = 0,\ \theta = 0)$, is sufficiently*

high,[15] *the audience also increases their belief that the leader took covert action.*

As suggested in the paper's opening anecdote, the public action itself can draw attention to the issue, and make rational audiences more suspicious that covert action is also being pursued outside of the public view (or at the very least, more suspicious that the leader is the kind of leader who *would* use covert action).

The value of using a cover story thus derives from reputational concerns resulting from the policy *outcome*, rather than just the actions taken. If the leader were confident that covert action would not succeed, there would be nothing for the leader to "cover up" (beyond the risk of direct exposure). It is the risk of policy success, and the need to provide some explanation for how that success came about, which drives the leader to employ a cover story.

Finally, we can examine more closely the conditions under which a cover story may be used in equilibrium:

Remark Cover stories are used with positive probability if $\lambda < \lambda' = 1 - \underline{k_p}/(\alpha_c \bar{r})$. This is more likely to be satisfied if:

- Transparency (λ) is low.
- The direct cost of public action ($\underline{k_p}$) is low.
- The effectiveness of covert action (α_c) is high.
- The leader's reputational concern ($\bar{r}$) is high.

This follows intuitively from the preceding discussion. The use of a cover story involves incurring a direct cost of ineffective public action ($\underline{k_p}$), in exchange for a reputational benefit ($\bar{r}$) which is enjoyed only if covert action is both successful (with probability α_c) and not directly exposed (with probability $1 - \lambda$).

2.5 Extensions

The models analyzed above made a number of simplifying assumptions, for the purpose of isolating the theoretical mechanisms of interest. Here we briefly demonstrate the robustness of the

[15] Specifically, if $Pr(a = (1,1) \mid \omega = 0,\ \theta = 0) > \frac{\tau_1}{\pi + (1-\pi)\tau_1}$

previous findings to two model extensions, which incorporate additional strategic considerations and add to the models' descriptive realism.

First, the preceding analysis demonstrated how the leader's use of cover stories can contribute to the audience's skepticism upon observing public action being taken, as per Corollary 2. This effect arose entirely through the indirect channel of the audience's beliefs regarding the leader's strategic behavior. However, we might expect that public action may also raise the risk of exposure of covert action more directly. For instance, public action may attract attention and scrutiny from the media or civil society actors; or it may place the target of the intervention on high alert, prompting the target to seek out direct evidence to implicate the Intervener.

We can incorporate this consideration into the model with the following extension. Suppose that the risk of direct revelation of covert action, $z \in \{0, 1\}$, is given by

$$Pr(z = 1|a) = a_c(\lambda + a_p\delta) \tag{5}$$

If the leader takes covert action alone, A observes $z = 1$ with probability $\lambda \in (0, 1)$; but if the leader also takes public action, the risk of exposure is increased to $\lambda + \delta$ for some $\delta > 0$.

With this modification of the model's setup, we have the following result:

Corollary 3 *Consider a variation of the heterogeneous-leader model, in which the revelation probability is given by (5). Then, for* $\lambda \leq \tilde{\lambda} := 1 - \left(\frac{k_p + \delta\bar{r}}{\alpha_c\bar{r}}\right)$*, there exists an equilibrium in which:*

- *L plays covert action with the same probability as characterized in the low-*λ *(*$\lambda < \lambda'$*) case of Proposition 2, that is,* $Pr(a_c) = (1 - \tau_1)F(k_c^\dagger)$*; and*
- *For some* $\bar{\delta} > 0$*, the probability that L plays a cover story is increasing in* δ *over the interval* $\delta \in [0, \bar{\delta}]$*.*

This corollary shows us that despite introducing an additional technological hurdle to the use of cover stories—in addition to the direct cost $\underline{k_p}$ associated with taking ineffective public action—the leader can still find it advantageous to use them, because of their impact on the audience's beliefs. Further, we see that, counterintuitively, the leader's equilibrium probability of using cover stories can actually *increase* in the additional scrutiny that they invite. This is due to the fact that, as δ

increases, the *absence* of direct evidence of covert action makes the audience increasingly confident that covert action was *not* used. The leader is then incentivized to take advantage of the audience's more favorable inference, by using cover stories more often.

As a second extension, we can consider modifying the policy production technology, as given in (1). The preceding analysis assumed, for simplicity, that when effective public action is being taken, covert action could not contribute to the probability of policy success: that is,

$$Pr(y = 1|a_p\omega = 1, a_c = 1) = Pr(y = 1|a_p\omega = 1, a_c = 0) = \alpha_p$$

Instead, let us now suppose that both policy levers can contribute to policy success:

$$Pr(y = 1|a_p\omega = 1, a_c = 1) = \alpha_p + \varepsilon, \quad \varepsilon > 0 \tag{6}$$

With this alternative setup, we have the following result:

Corollary 4 *Consider a variation of the heterogeneous-leader model, in which the policy production function satisfies (6). If ε is not too great, then for $\lambda < \lambda'$, there exists an equilibrium in which either:*

(i) The path-of-play behavior is identical to that of the equilibrium characterized in Proposition 2;

Or:

(ii) A plays the same strategy as characterized in Proposition 2; L takes covert and public action simultaneously with positive probability when $\omega = 1$; and the probability of L using a cover story is increasing in ε.

Recall that a cover story was defined as the use of *ineffective* public action, when $\omega = 0$, alongside covert action. It is unsurprising that this modification of the model setup increases the simultaneous use of public and covert action when $\omega = 1$, which is the state in which the probability of policy success is directly altered. Less intuitive is the fact that it also does so when $\omega = 0$. The reasoning behind this result is roughly similar to the previous one: the audience's inference upon observing public action and no direct revelation of covert action (and specifically, when the policy

does not succeed) becomes increasingly favorable as ε increases; so the leader is incentivized to exploit this situation by playing a cover story with higher probability.

2.6 Empirical Predictions

Our formal analysis yields several novel empirical implications for studying the decision to authorize secret policies, such as covert intervention. First, we should observe that decision-makers within Intervener states exhibit a concern for strategic inferences formed by the audiences among whom they hope to maintain a positive reputation. That is, when planning covert actions, decision-makers will not only consider the operational security and the risk of direct exposure of those actions; they will also consider how they are perceived by a skeptical audience—even in the best-case scenario that the operation succeeds with no direct exposure—and how they might be able to allay the audience's suspicion of their involvement. Recall that our general model places no restrictions on who the audience is; we only assume that the audience has an interest in preventing the leader from engaging in unscrupulous covert action, or in punishing the kinds of leaders who do so.

Second, when the level of transparency (λ) around a foreign policy issue is low, the Intervener will use overt action as a cover story for their covert intervention. The overt action should be a lesser violation of international laws and norms, and thus less objectionable to the audience, as compared to 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 can significantly contribute to the likelihood of policy success ($\alpha_p > \alpha_0$).

Third, the Intervener should make an effort to connect the favorable policy outcome to the public action in the mind of the audience. Evidence on this point could include public statements attempting to draw such a connection, or even more convincingly, private documentation of the Intervener's intent to appeal to public actions as a mechanism to disclaim responsibility for secret policies. Finally, the audience(s) of interest to the Intervener should be convinced by the cover story, and willing to attribute the observed outcome, in whole or in part, to the observed public actions.

We examine each of these implications in the case study that follows.

3 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 in Guatemala (Fraser, 2005, 487). But it was not working for the United States. Answering the calls of the Guatemalan communist party, newly-elected President Jacobo Arbenz implemented extensive land and agrarian reforms (Schlesinger and Kinzer, 1982, 53), which directly challenged U.S. commercial and political interests. U.S. policymakers were also concerned by the number of communists appointed to government positions (Immerman, 1982, 108). In his memoirs, Eisenhower worried that a major threat to his objectives was that "Communism was striving to establish its first beachhead in the Americas by gaining control of Guatemala."[16]

In August, 1953, Eisenhower authorized the covert CIA operation PBSUCCESS. The first phase of the operation 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 real 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 aimed at convincing loyalists that defense of Arbenz was futile and would lead to reprisals. This included a media blitz across Latin America, bribes to Guatemalan politicians to have them recognize the coup plotters as the rightful governments, and threats against those whom they could not buy (Schlesinger and Kinzer, 1982, 114). The paramilitary operations only commenced after months of psychological operations had already begun to undermine widespread confidence in the Arbenz government.

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 U.S. retained enough plausibly deniability to avoid backlash.

Following best practices in the evaluation of formal models, we use this case to illustrate the empirical plausibility of our theory. In section 3.1 we detail our case selection methodology (Bates,

[16] Quoted in Schmitz (1999, 179)

1998). Then, following Goemans and Spaniel (2016), Joseph, Poznansky and Spaniel (2022) and others, we evaluate our theory by examining primary evidence of the Eisenhower administration's decision-making processes, paying particular attention to the choice nodes that we model. We develop case-specific hypotheses about what our theory predicts we should observe, and evaluate them against the leading alternatives. Section 3.2 demonstrates that the Eisenhower administration worried about strategic inferences as they implemented their covert actions. Section 3.3 details the overt actions the administration took, and explores how well existing accounts can explain these actions. Section 3.4 provides evidence that the cover story mechanism can explain these overt actions. Finally, section 3.5 addresses concerns and alternative explanations.

3.1 Case selection, and calibrating the parameters

Following Bates (1998), our main concern was finding a case in which the leader of the intervening state faced the core strategic tension characterized by our model, and the initial conditions fit the parameters that support the cover story equilibrium. Below, we first discuss why the Eisenhower administration's overarching objective of toppling communist-leaning but popular governments fits the broad contours of our model. We then discuss why the Guatemala case in particular closely matches the conditions for the cover story equilibrium.

3.1.1 Selection of the Eisenhower administration

Our theory applies to a major world power that is both interested in shaping political developments abroad, and concerned with maintaining a reputation among various audiences for only doing so through scrupulous means. This makes the United States in the latter half of the 20$^{\text{th}}$ century a natural choice. The early Cold War, and the Eisenhower administration in particular, are especially appropriate. Eisenhower's primary foreign policy objective was to stop the spread of communism (Schmitz, 1999). To win over the developing world, the U.S.'s overarching strategy was to promote the principles of sovereignty, self-determination and democracy as core tenants of the Liberal Order (Rabe, 1988, 166). Yet there remained uncertainty across the developing world as to the U.S.'s true commitment to the ideals it espoused—reflecting uncertainty regarding the leader's true "type", as analyzed in our adverse selection model. Eisenhower understood that using military power to overturn a democratically elected government would reveal him as highly

unscrupulous, in the sense implied by our theory (Poznansky, 2019, 86). In the Guatemala case specifically, the Administration estimated that such overt disregard for liberal principles would "stigmatize our international reputation."[17] Thus the core tension of wanting to shape political developments abroad, while avoiding the reputational damage that would follow from doing so through unscrupulous means, is a prominent tension faced by the Eisenhower administration in this early Cold War period.

One concern with selecting the Eisenhower administration arises because some question the extent of President Eisenhower's direct involvement in foreign policy choices, and the degree to which administration policies actually reflected his own worldview (Divine, 1981). This work instead suggests that key advisers, notably the Dulles Brothers (with Allen Dulles as CIA director, and John Foster Dulles as Secretary of State), played an outsize role. Yet other work argues Eisenhower was more skillful and directly involved in policy decisions (see McAuliffe, 1981). Recognizing that this debate exists, we analyze documents that provide insights into the reasoning of the Administration as a whole—including Eisenhower and the Dulles brothers, as well as their subordinates within the CIA's Directorate of Plans and the Guatemalan Embassy.

Another concern is that the CIA was unusually popular with the U.S. public in 1954, and that therefore the U.S. public would have ignored even direct evidence of a covert operation. However, support was unusually strong only because "American people remained in blissful ignorance of the CIA's covert objectives" (Jeffreys-Jones, 2022). If those actions were exposed, opinions may have changed. Further, in the case we analyze, administration officials primarily expressed concern over international audiences (including audiences within Guatemala, as well as throughout Latin America and beyond) rather than U.S. domestic audiences when discussing the risk of exposure for PBSUCCESS (Schmitz, 1999). The relative concern over different audiences may shift across different cases and different time periods, and we believe our theory can accommodate this variation.

3.1.2 Selection of the Guatemala intervention

Guatemala is an especially important case to examine (Bates (1998)'s second criterion) because it represented the first major communist foothold in the Americas. The analytical clarity of this particular case, relative to other regime change operations that Eisenhower authorized and

[17] See Memorandum for Col J. C. King, PBSUCCESS 20th Jan 1954.

pursued, is also aided by the fact that all the salient choices were made within the Eisenhower administration.[18]

Consistent with our model, in pursuing regime change in Guatemala, the Eisenhower administration faced policy options that can largely be characterized as either public and scrupulous, or covert and unscrupulous. Concerned that brazen military intervention into a regional democracy would sour opinions of the U.S. throughout Latin America (Schmitz, 1999, p181), Eisenhower only seriously considered military actions that could be undertaken covertly. By contrast, economic sanctions, or public diplomacy that was designed to expose the failures of communism and cause domestic unrest, were not seen as inconsistent with Liberal Order, and thus more tolerable to foreign and domestic audiences.[19]

As discussed in the formal analysis, the necessary conditions for the cover story equilibrium to hold are that the transparency parameter, λ, is low; the effectiveness of covert action, α_c, is high; the leader's reputational concern, $\bar{r}$, is relatively high; and the direct cost of public action, $\underline{k_p}$, is moderately low, but not zero. At this time, CIA was optimistic that they could successfully orchestrate a coup (high α_c) and conceal direct evidence of their involvement (low λ). Their confidence was heightened after using the same playbook to oust Mossadeq in Iran undetected (Cullather, 2006, 7). The state of information technology in Guatemala in 1954 also meant the chance of direct exposure was low (Joseph and Poznansky, 2018). Further, in this period, it was not widely known whether the CIA was even in the business of orchestrating regime change, or if so, by what methods. Thus even if some audiences sought out direct evidence of U.S. covert involvement, they may not have even known what to look for. The preceding discussion of the U.S.'s concern for its reputation across the developing world, along with the fact that Eisenhower was facing reelection at home, imply a high value of $\bar{r}$. Finally, we demonstrate below that the public actions pursued as a cover story involved direct costs that were not negligible, but were substantially outweighed by reputational considerations ($\underline{k_p}$ moderately low).

[18] As another potential case, many features of Eisenhower's reasoning to oust Castro fit our cover story equilibrium. But the case is more complicated to analyze because the decisions spanned multiple administrations, with Kennedy ultimately approving the mission. For an interesting overview of cover story references in this case, see CIA, Official History of the Bay of Pigs Operation, V II, pp12-14.

[19] See Memorandum for Col J. C. King, PBSUCCESS 20th Jan 1954.

3.2 Plausible deniability was difficult but important

From the outset of planning Operation PBSUCCESS, plausible deniability was viewed as essential to the mission's success. A recurring reminder from administration officials to mission planners was: "don't get caught" (FRUS, 1954). Consistent with existing theoretical arguments (Joseph and Poznansky, 2018), this included diligent efforts to avoid direct evidence of U.S. involvement. According to Immerman (1982, p133) "Planning took place with the utmost stealth. Only Eisenhower, the Dulles brothers, and a few other top-level members of the White House, State Department, and Central Intelligence Agency knew that an operation was even being considered, let alone were privy to its details."

But there is only so much that careful planning can do. With the stationing officers 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. After the active phase of PBSUCCESS was given the "full green light" in April 1954, CIA officers remained in Guatemala and South America to facilitate psychological operations, bribe Guatemalan politicians and military officers, and otherwise monitor the plot (Cullather, 2006).

3.3 The Puzzle of Overt Action

Given the intense focus on maintaining secrecy, we might expect that the administration would aim to minimize the risk of direct exposure by diverting public attention away from Guatemala as CIA officers were in the field. But this is not what we obseve. In this section, we describe a series of public actions that the administration took, and discuss why they are not consistent with a simple alternative explanation of using all available and effective means to advance the policy objective. In the next section, we discuss audience perceptions of the efficacy of these public actions.

The U.S.'s diplomatic posture in the lead up to the coup certainly gave no impression of U.S. disinterest in political developments within Guatemala. In early 1954, U.S. Ambassador to Guatemala John Peurifoy and others made inflammatory statements that the U.S. 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 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 a secret, but the U.S. discovered it the next day (Cullather, 2006, 80). Rather than minimize the episode, Eisenhower expressed public outrage. He invoked the Monroe Doctrine, which called for the exclusive influence of the United States in Latin America, and proceeded to impose a naval blockade to prevent future arms shipments into Guatemala (Cullather, 2006, 79). In fact, from the U.S. 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 U.S. would then "discover" and exploit publicly. The convenient occurrence of an actual weapons shipment obviated the need for this particular ploy.

Around the same time, the U.S. convened an emergency meeting of the Organization of American States in which Dulles delivered an impassioned speech attacking the Guatemala 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 U.S." By "proper", Eisenhower meant public and short of calling for military intervention (Bowen, 1983). After months of delay, the Executive also authorized a Memorandum of Understanding with Honduras on military exchange, with the goal of enhancing protection from neighboring communist states (i.e. Guatemala).

Why would Eisenhower shine a light on Guatemala when covert operations were underway? The conventional explanation is that mission planners wanted to maximize the chance Arbenz would step down, by ramping up psychological pressure and weakening his capacity to resist the paramilitary operations. This led Eisenhower to authorize all available policies, both overt (but short of military intervention) and covert (e.g. Cullather, 2006, p59).

This argument is not inconsistent with our theory, which allows for the possibility that some public actions are effective and are undertaken for that reason. The important question for our analysis is: does Eisenhower's desire to deploy all effective policy instruments fully account for the overt action that we observe? If the answer is yes, we would expect the administration to only

publicize overt policies when doing so confers some operational advantage. We believe that two aspects of how Eisenhower publicized overt actions are incompatible with this explanation.

First, the executive publicized events within the United States. In fact, DCI Dulles deliberately exaggerated the scope of the weapons shipment to prompt Congressional statements and press coverage (Cullather, 2006, p59). There were operational disadvantages to engaging the U.S. public directly. One concern was that PBSUCCESS was commanding operations from an undisclosed location in Florida (codenamed LINCOLN). The more attention within the United States, the more media scrutiny would follow, raising the chance of exposure at this critical operational moment. Further, Assistant Secretary of State Cabot had previously warned that if U.S. "public opinion should become too aroused and excited, there might be embarrassing demands for [overt] action... [that were] altogether infeasible" (CIA, 1953).

Second, while PBSUCCESS relied partly on broadcasting anti-Arbenz messages across Guatemala, mission success did not rely on messages voiced from 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*h*). So it is not clear why Eisenhower would call on diplomatic staff to directly voice anti-Guatemalan positions when PBSUCCESS was operating local radio stations that could have voiced the same messages.

3.4 The cover story explanation

As outlined above, our theoretical model carries four key observable implications. First, actors within the administration will exhibit concern for strategic inferences made by audiences, even in the absence of direct evidence of wrongdoing. Second, the administration will pursue overt, performative policies ostensibly targeted toward the same objective being pursued simultaneously through covert means. Third, those privy to operation PBSUCCESS will attempt to convince audiences that no covert action took place, referencing the public actions taken as an alternative explanation. Finally, despite the audience's understanding of U.S. interests and capabilities, they will not be suspicious enough of U.S. covert action to demand punishment or retribution; rather, they will be convinced, to a sufficient degree of confidence, that the succesesful outcome is attributable to

the observed public actions. Evidence of the second implication (public actions taken alongside PBSUCCESS) was discussed in the previous section. Below we consider each of the remaining implications in turn.

3.4.1 Concern for strategic inferences

While planning PBSUCCESS, administration officials expressed an acute concern for strategic inferences. The NSC explicitly acknowledged that even if no direct evidence of CIA involvement was revealed, "It must be recognized that any major effort to dislodge the Communist-controlled government of Guatemala will probably be credited to the United States, and possibly on CIA." As a result, "Covert accomplishment of the objectives of PBSUCCESS is therefore defined as meaning accomplishment with plausible denial of United States or CIA participation" (FRUS, 1953) after the operation was concluded. Consistent with our theory, the NSC defined success in terms of overall perceptions of U.S. involvement, even absent direct evidence.

CIA Deputy Director for Plans Frank Wisner laid out the concern even more explicitly. Wisner cautioned 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*f*). 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 U.S. But it is equally true that they would see the hand of the U.S. in any uprising whether or not sponsored by the U.S." (CIA, 1954*e*). By the logic of our theory, Wisner is articulating the view that, given a low level of transparency (low λ), the *absence of evidence* of U.S. involvement does not provide sufficiently compelling *evidence of absence* of U.S. involvement in order to avoid blame for the observed outcome.

Given these concerns, Wisner and his staff took an active role in crafting cover stories to allay suspicions among the target audience. In a discussion about how to prevent Latin American audiences from speculating of U.S. involvement, Wisner argued that "it might be a good idea to cry wolf several times before D-Day" (CIA, 1954*g*). In June, Wisner's subordinates managing operations from LINCOLN observed with disappointment that U.S. Ambassadors in Honduras and Guatemala were not publicly voicing the U.S. position. Even though they did not think diplomatic statements would affect whether or not the coup prevailed, they still thought it was "essential that

for diplomatic battle the hole created by non-participation should be filled" (CIA, 1954*d*).

3.4.2 Employing the cover story

We expect that administration officials should seek to explicitly connect the observed outcomes to the public actions, so as to disclaim any covert regime change operations. This is a hard test, because political elites rarely document the reasoning behind what they say. However, elites privy to the operation explicitly documented a cover-story motivation for their discussion of U.S. overt policies as PBSUCCESS was underway. For example, Second Secretary Hill of the U.S. Embassy in Guatemala recounted his conversation with a Guatemalan political elite (whose name is still classified)[20] as follows:

> I told [redacted] that Ambassador Patterson had been quite correct in pointing out the U.S. policy of non-intervention... but [redacted] was quite wrong in thinking that the U.S. 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 U.S. had no concrete plan for intervention in the domestic affairs of Guatemala and continued its non-intervention policy.* (emphasis added)[21]

This last sentence directly describes the logic of our argument: the reason that Hill highlights overt policies is to disclaim involvement in covert policies.[22]

Furthermore, after Arbenz fell, we find that U.S. officials continued to publicize the overt actions taken in order to cover up the covert actions as the public, media and international community wondered about their involvment. An NSC report, later released to the press, read:

> The Organization of American States was used as a means of achieving our objectives in the case of communist intervention in Guatemala. After the arrival from Poland

[20] The context suggests the unknown subject was influential in Guatemalan politics, not currently in government, and somewhat concerned about the communist trajectory, but not loyal to the United States.

[21] CIA (1954*b*)

[22] For a second example with a different unknown subject, see CIA (1954*a*).

> on May 15 in Guatemala of a substantial shipment of arms, the United States initiated consultations with all Latin American Governments, except Guatemala. Following these consultations, the Council of the Organization of American States voted almost unanimously... to convoke a Meeting of Ministers of Foreign Affairs. The Council of the OAS postponed the meeting sine die because the revolution in Guatemala overthrew the communist-controlled Government.
>
> The revolution in Guatemala caused the communist-controlled Arbenz Government to appeal to the UNSC and to the Inter-American Peace Committee of the OAS alleging aggression on the part of Honduras and Nicaragua, supported by other foreign nations. The United States took the position that the Organization of American States was ready, willing and competent to respond to the appeal... The Inter-American Peace Committee prepared to investigate, but before the Committee arrived in Guatemala, the new government of that country indicated that the controversy requiring the investigation had ceased to exist (CIA, 1955).

As the quote shows, the story being put forward by the administration is that the U.S. pursued its objectives in Guatamala through the OAS; by implication, the U.S. would not have been involved in other, unscrupulous means of pursuing the same objectives.

3.4.3 Public reception of the cover story

Finally, while we expect that observers will be naturally suspicious of a U.S. covert action, their suspicions will be offset in part by the lack of direct evidence, and in part by their belief the U.S. public actions contributed to ousting Arbenz. Crucial to our argument is that not only did the U.S. attempt to link the observed outcome to the public actions they took, but their attempts were largely successful.

Several journalists and academics at the time analyzed the extent and impact of U.S. involvement in the Guatemala Affair. Notably, two years after Arbenz was ousted, Taylor (1956) published a comprehensive "Critique of United States Foreign Policy" surrounding Arbenz's removal. He reviewed journalistic inquiries into U.S. policies, and academic and policy investigations into the U.S. role published across Latin America and the United States. He also relied on interviews with those

in Guatemala and Honduras at the time.

Considering the question of whether the U.S. was directly involved in plotting the coup or training its perpetrators, Taylor reaches the following determination:

> 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 U.S. arms exports reached the revolutionaries indirectly, by way of third countries friendly to the U.S. On this point, he finds:

> The conclusion that the United States played an important part in the struggle in Guatemala seems inescapable. 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 Armed forces. ... But it can be shown that the United States played a role in the United Nations which tended to deny to Guatemala the privileges apparently guaranteed it by its membership in that organization. (p. 797)

Given these observations, he asserts conclusively: "The inaction of the U.N. Security Council and of the Inter-American Peace Committee (as agent for the O.A.S.) 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 U.S. did not directly contribute to the revolt relies on two premises: first, that no direct evidence exists; and second, that the U.S. was taking meaningful (and publicly observable) actions that tilted the balance in Castillo Armas' favor. In this case, his argument is that U.S. overt actions contributed by denying Arbenz the option to appeal to regional security institutions for support, which altered Arbenz' incentives to step down; and that U.S. public statements served to encourage and embolden Arbenz' opponents into mobilizing against him.[23] He later supplements his argument by stating that U.S. blockades denied Arbenz weapons and reduced the probability that resistance to the coup attempt would

[23]This latter point is consistent with recent theoretical work demonstrating that international messages can facilitate local actors in coordinating for regime change (Little, 2017; Malis and Smith, 2019).

have been successful, again speaking to the perceived efficacy of U.S. overt policies (p804-805). Elsewhere, he argues that U.S. open support encouraged OAS members hostile to Guatemala to facilitate the revolution. Specifically, he asserts that Honduras failed to meet their OAS obligation by allowing the rebels to board a plane to Guatemala, and that had the U.S. not so brazenly opposed Arbenz, Honduras may have acted differently.

Taylor's account is compelling because it is directly focused on evaluating the efficacy of U.S. open policies, and the chance of covert policies based on a wide range of sources. What is more, Taylor was an established U.S.-based historian of Latin America, with no direct government affiliation and who was partly sympathetic to socialist movements in other work (e.g, Taylor, 1967). Therefore, he is likely to provide a somewhat balanced perspective. But similar arguments, all consistent with our story, are presented from a broader range of journalists. For example, we observe the reasoning we predict from the more pro-U.S. Harsch (1954), who acknowledged U.S. incentives for covert action, stating "If there were no native revolutionary movement to encourage and support, then some other ... remedy would have to be found." But he does not believe that covert action was necessary here 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. We even observe our logic in cases where reporters are deeply critical of U.S. strategy in Latin America, and deeply suspicious that Dulles will openly meddle in Latin America. For example, Reston (1954) openly speculates that the CIA was involved in Guatemala, but stops short of asserting that they were directly involved in the coup—instead noting that the CIA was integral in uncovering the weapons cache, and exploiting that episode to foster anti-Arbenz sentiment.

The broader reactions at the time are consistent with our theory in two other ways. First, our theory assumes that overt actions are costly, but less costly than covert actions. The U.S. was broadly criticized for overt actions that were viewed as intrusive meddling against a democratic elected government. For example, several Latin American states viewed the blockade as an unjustified violation of sovereignty; but the backlash was relatively minor (Friedman, 2010, 672), particularly compared to the backlash that would have followed from the revelation of the U.S.'s even more unscrupulous covert activities.

Second, some pundits speculated about U.S. involvement shortly after Arbenz fell because they understood U.S. incentives were consistent with the outcome (e.g Grant, 1955). Thus, Eisenhower did not completely escape strategic inferences. This is supportive of our mechanism. In our theory, public actions do not entirely prevent strategic inferences; rather, they offset suspicion enough for the Intervener to avoid backlash. Consistent with our theory, some suspicion did arise. But that suspicion was not widespread. Indeed, historians who consider sustaining plausible deniability as part of mission success, argue that PBSUCCESS was a successful case of covert regime change (Immerman, 1982; Schmitz, 1999).

Altogether, this evidence shows that before the mission was carried out, the executive was concerned about strategic inferences; and that as part of the mission planning, the CIA conceptualized a diversionary public action so that they could retain plausible deniability in the face of these strategic inferences. After the mission was carried out the administration referred back to these overt policies to divert attention away from their sponsorship in the years after the coup succeeded. Further, analysts at the time estimated that U.S. overt actions meaningfully contributed to the mission success, and partly used those arguments to determine that it was unlikely that the U.S. was directly involved in planning the revolution.

3.5 Clarifications

Following Bates (1998) and others, we narrowly focused on the aspects of the case that are most salient for the decision nodes in our theory. We now address two potential concerns that might be raised by researchers who study PBSUCCESS from a broader range of historical perspectives.

First, while our goal was to show that the cover story mechanism was one overlooked way that mission planners avoided strategic inferences, we recognize that it was not the only way. The CIA deliberately trained Guatemalan exiles to make the coup appear like a local conflict between Guatemalan political factions. The CIA also crafted the appearance of alternative foreign sponsors; most notably, the CIA deliberately trained and armed the Guatemalan coup plotters in Nicaragua, Honduras and other countries that were hostile to Arbenz. Of course, training forces overseas raises the risk of direct exposure because the CIA cannot easily control the environment.[24] The CIA also

[24]There was a near miss in January 1954, when chatter from Nicaraguans privy to local operations prompted Guatemala to published a White Paper accusing "the government of the North" of supporting covert, anti-Guatemalan activities in Nicaragua. However, the chatter was unsubstantiated, and could have referred to Mexico. According

armed the coup-plotters with weapons it purchased from the Dominican Republic to implicate them (CIA, 1954*c*). We do not think plausible deniability is uni-causal, and we view these methods as consistent with our overall theory. After all, each method raised the risk of direct exposure against the benefits from reducing strategic inferences. As the Operations Coordinating Board put it in a Memo designed to assess plausible deniability, "Added support in cloaking the U.S. hand exists in the number of other countries"—such as Nicaragua and Honduras, where the foreign training bases were located—"which both have good reasons for wanting to see the replacement of the Arbenz Government and have the means for backing a coup of the size planned."[25]

Second, one might wonder if Eisenhower engaged others at the OAS to offset backlash in the event that U.S.actions became public. This would not be inconsistent with our argument if this objective followed alongside the cover-story objective. However, it is notable that we found evidence of the cover story mechanism in NSC deliberations, and exchanges between Eisenhower and Dulles. We did not find any discussion of gaining consensus in the case that the covert action was exposed. It is also worth noting that this logic could not explain Eisenhower's choice to publicise the blockade or make inflammatory statements against Arbenz outside of the OAS meetings.

4 Broadening the Argument

Our proposed cover story mechanism could apply in any scenario in which a government has multiple policy levers, one of which is undesirable to an important audience (foreign or domestic), and the government expects that it may be able to conceal its use of the undesirable lever. As summarized in Table 1, we now present two additional vignettes to demonstrate that this scenario is reasonably general. While not claiming to provide a comprehensive explanation for either case, we aim to demonstrate that our mechanism could plausibly illuminate underappreciated dynamics that arise across diverse policy scenarios of interest to a broader range of scholars.

to the CIA, "Continued study of the aftereffects of the White Paper indicates that it somewhat reinforced suspicions among all those previously inclined to suspect the U.S. but was roundly disbelieved by the majority of anti-Communists in Central America."

[25] https://history.state.gov/historicaldocuments/frus1952-54Guat/d133

Model	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 exchange (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 interest of politically connected Australian firms at risk of national reputation	Appearing weak to electorate, creating moral hazard, NATO repercussions

Table 1: Summary of diverse features of cases

4.1 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 re-negotiate the existing oil concessions, which heavily favored Australia (Australia, 2000). This created a vexing policy challenge for the Australian Government (AG). On the one hand, AG viewed Timor Gap profits as an important national interest because they generated enormous tax revenue and high-paying jobs.[26] The AG also faced political pressure from large, politically connected Australian firms that operated the oil concessions at the time. Further, experts were concerned with follow-on effects because "Indonesia will feel quite aggrieved if we have unequal boundaries in certain areas with Indonesia and we suddenly blow the boundary out and make a more equidistant one in relation to East Timor (Pugh, 2000)." However, several legal analysts believed that East Timor could accrue substantial concessions if the matter was referred to an international court (King, 2017).

On the other hand, the international community, especially the United States, supported fairer terms. At minimum, they expected Australia to negotiate fairly. US ambassador Peter Galbraith

[26] As Dodd (2007) argues, it was so important that Foreign Minister Alexander Downer invoked a 'national interest' exemption clause to fast-track ratification of CMATS treaty without scrutiny by the Joint Standing Committee on Treaties. This exemption has been used only six times in Australia's history.

was appointed to negotiated on behalf of East Timor. As negotiations were ongoing, over 50 U.S. members of congress, including Nancy Pelosi, Jack Reed, and Patrick Leahy, wrote the Australian Prime Minister calling on Australia to adhere to strict legal principles during the negotiations (Frank, 2004). East Timor was ranked below 160th of all nations in terms of political and economic development. It also stood out "as the most oil-dependent country in the world. [Even] In 2009, petroleum income accounted for about 95 percent of total government revenue and almost 80 percent of gross national income (IMF, 2011)." These American elites, who had supported East Timor's independence at the cost of tension with Indonesia, worried that absent the Timor Gap revenue, East Timor would devolve into a failed state. Another issue for Australia 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 border dispute (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. Indeed, 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 in the negotiation outcome. 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). AG also stalled profit sharing between 2003 and 2004, demanding that East Timor make important concessions. This provoked backlash from American elites, who argued that Australie was taking advantage of their neighbor's impoverished position (Frank, 2004). Still, in 2006, Amb. Galbraith and East Timor accepted CMATS, believing that Australia had adhered to international law during the negotiation.

This was not the case. Secretly, the Australian Secret Intelligence Service, who were invited into East Timor as part of a counter-terrorism operation, illegally bugged the Office of East Timor's president and other key negotiators (Cannane, 2015). Thus, unbeknownst to Amb. Galbraith and

the East Timorese, the Australian Government new exactly what the negotiators were willing to accept. Australia's secret efforts were publicized by a national security 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 US reaction is consistent with our characterization that US elites wanted to prevent objectionable means independently 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 whole experience of the negotiation from 2000 on and through this whole episode was to see a country that—yes, in many ways focuses on the public good—but where corporate greed was a big part of it, because 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.

4.2 Cuban Missile Crisis

Officially, 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 ended because Kennedy secret committed 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[27] should the quid pro quo become public (Bernstein, 1980). For example, instead of delivering a letter from Khrushchev to President Kennedy, Robert Kennedy

[27] Air Force Major General William Seneter (1963), for instance, expressed concern that Turkey would doubt the US security guarantee if they that discovered the US traded away the Jupiter missiles as part of the exchange.

returned the letter to the Soviet Ambassador explaining, "I myself, for example, do not want to risk getting involved in the transmission of this sort of letter, since who knows where and when such letters can surface or be somehow published... The appearance of such a document could cause irreparable harm to my political career in the future" (FRUS, 1962).

Our theory sheds light on two underappreciated aspects of this case. First, 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 US missiles were close to the Soviet Union to justify the Soviet missile deployment to Cuba. Even at the time, many speculated that a missile exchange could have facilitated peace.[28]

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). Shortly after the crisis, Kennedy told Eisenhower and Truman in private conversations that he did not exchange missile removals. Because these conversations were in confidence between old friends, they had the air of credibility when they eventually leaked. Kennedy also vilified UN Ambassador Adlai Stevenson, who was the sole advisor to advocate for missile exchange during the crisis. Finally, Kennedy minimized suspicion by waiting five months to remove the missiles from Turkey, and by removing missiles from Italy and Turkey at the same time, to make it appear as if it was part of a broader effort to restructure forces.

Our theory suggests that the non-invasion pledge played a more important role in advancing this fiction than scholars appreciate, because it gave Kennedy some position to fall back on. 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). This would not have been possible without an alternative position. To be clear, we accept that the non-invasion pledge did not offset all suspicion. By our theory, the goal is not to offset all suspicion; rather, the goal is to offset enough suspicion to avoid backlash. Indeed, the amount of suspicion that was raised did not encumber Robert Kennedy from advancing to the Senate, and from running for president, as he feared a public disclosure of this episode would.

A second aspect of the case illuminated by our theory is the *indirect* effect of audience costs

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

(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 (Kurizaki and Whang, 2015). But in this case, Kennedy made public statements that promised the US would not back down, while he secretly offered the Soviets a substantial concession. This raises the question: why did Kennedy make these public statements at all?

We argue that audience costs could raise the credibility of 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."

5 Conclusion

We argued that when international and domestic audiences form perceptions and understandings of international affairs, they draw inferences about the actions of powerful states not only from direct evidence, but also from circumstantial evidence—including context, interests, capabilities, and policy outcomes. This insight illuminates a core tension states face if they hope to sustain plausible deniability for their secret policies. On the one hand, they must avoid getting caught as they plan and execute their operations. On the other, after they succeed, they need to explain how the world turned in their favor by chance, and not as the result of a repugnant action they took in secret. How can they simultaneously achieve these goals?

When policy success is unlikely to occur in the absence of intervention, a state can avoid attribution of their covert action through the use of a cover story. The ideal cover story is more consistent with international laws and norms than the secret policy that the leader wants to disclaim responsibility for, and constitutes a plausible explanation for observed policy outcomes. Cover stories are rational mechanisms to deny morally repugnant, illegal or otherwise objectionable secret policies, even if they invite additional scrutiny that raises the risk that such policies are exposed directly, and even if they are somewhat unpopular among international or domestic audiences. In fact, the more scrutiny they invite, the more effective they are at disclaiming secret policies ex-post.

Planning that surrounds Operation PBSUCCESS, the CIA mission to remove Arbenz as president of Guatemala, supports our theory. We find evidence that the Eisenhower Administration worried about strategic inferences, believing that many in the region would assume the US was responsible even if no direct evidence emerged of CIA involvement. They deployed a range of tactics to avoid attribution for Arbenz's ultimate ouster. To this end, one important but overlooked mechanism was a series of performative overt policies. We find direct evidence that Embassy staff in Guatemala discussed US public actions to disclaim the possibility of covert regime change operations. We also show that in the years after Arbenz resigns, the administration points to these open policies to account for its own involvement in Arbenz's removal, and to help discredit speculation that a nefarious secret policy was at work.

Our theory holds two primary policy implications. First, it helps public accountability activists appreciate 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. This, in turn, may make covert action more attractive.

Second, the logic of strategic inferences means that grey zone conflict may play a different role in great power competition—particularly, the emerging US-China competition—than existing research would predict. Existing work emphasizes that states can avoid escalation and retaliation if they can conceal direct evidence of an attack. But the United States often employs secrecy to maintain its reputation as compliant with liberal values, while pursuing goals that clearly violate those principles. To the extent that strategic inferences are plausible, grey zone attacks may not be viable; and this constraint may be more binding on the US, as the power most expressly concerned with maintaining the principles of the Liberal Order. This may present an autocratic advantage in the use of grey zone conflict to influence third parties in the decades to come. Our theory also shows that this disadvantage can be partially offset via a cover story.

References

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

ADST. 2023. "Oral Histories Project, Cuba.". Association for Diplomatic Studies and Training.

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.

Ashworth, Scott and Ethan Bueno de Mesquita. 2014. "Is voter competence good for voters?: Information, rationality, and democratic performance." *American Political Science Review* 108(3):565–587.

Ashworth, Scott and Kristopher W Ramsay. 2024. "The accountability of politicians in international crises and the nature of audience cost." *Political Science Research and Methods* 12(1):1–26.

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.

Barro, Robert J. 1973. "The control of politicians: an economic model." *Public choice* pp. 19–42.

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.
URL: *http://www.jstor.org/stable/2149587*

Biddle, Stephen, Julia Macdonald and Ryan Baker. 2018. "Small footprint, small payoff: The military effectiveness of security force assistance." *Journal of Strategic Studies* 41:89–142.

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

Bueno de Mesquita, Ethan Bueno. 2007. "Politics and the suboptimal provision of counterterror." *International Organization* 61(1):9–36.

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. 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*. "Conversation Between U.S. Ambassador Hill and SA/DD/P, 25 Jan (CREST no. 0000914014).". https://www.cia.gov/readingroom/docs/DOC_0000914014.pdf.

CIA. 1954*b*. "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*c*. "Memorandum for Goodbourne, Progress Report - PBSUCCESS. 11-17 May (CREST No. 0000923551).".

CIA. 1954*d*. "Memorandum from LINCOLN to RYBAT (CIA) Director, 21 June, 1954." *Central Intelligence Agency, Job 79-01025A, Box 9, Folder 2.* .

CIA. 1954*e*. "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.* .

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

CIA. 1954*g*. "PBSUCCESS Weekly Meeting, 30 March 1954 (Crest No. 0000916849).". https://www.cia.gov/readingroom/docs/DOC_0000916849.pdf.

CIA. 1954*h*. "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. 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.
URL: *http://doi.wiley.com/10.1111/j.1540-5907.2011.00567.x*

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.
URL: *https://books.google.com/books?id=eDsG-0-p15sC*

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

DoS. 2023*b*. "History of the U.S. and Bulgaria.". U.S Embassy in Bulgaria (Department of State).
URL: *https://bg.usembassy.gov/our-relationship/policy-history/io/*

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*

Farrell, Henry and Abraham L. Newman. 2021. "The Janus Face of the Liberal International Information Order: When Global Institutions Are Self-Undermining." *International Organization* 75:333–358.

Fearon, J. D. 1994. "Domestic Political Audiences and the Escalation of International Disputes." *American Political Science Review* 88:577–592.
URL: *http://journals.cambridge.org/abstract˙S0003055400093667*

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.

Ferejohn, John. 1986. "Incumbent performance and electoral control." *Public choice* pp. 5–25.

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

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

FRUS. 1962. "Telegram from Soviet Ambassador to the US Dobrynin to the USSR Foreign Ministry (Translation).".

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.

Hakakian, Roya. 2011. *Assassins of the Turquoise Palace*. Grove Press.

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

Hawkins, Darren, David A. Lake, Daniel L. Nielson and Michael J. Tierney. 2006. *Delegation under anarchy: states, international organizations, and principal-agent theory*. Cambridge University Press pp. 3–38.

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"..".

Huth, Paul K., Sarah E. Croco and Benjamin J. Appel. 2011. "Does International Law Promote the Peaceful Settlement of International Disputes?" *American Political Science Review* 105:415–436.

IMF. 2011. "Democratic Republic of Timor-Leste: 2010 Article IV Consultation—Staff Report.".

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

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 and Taehee Whang. 2015. "Detecting Audience Costs in International Disputes." *International Organization* 69:949–980.

Lake, David A., Lisa L. Martin and Thomas Risse. 2021. "Challenges to the Liberal Order: Reflections on International Organization." *International Organization* 75:225–257.

Levin, Dov H. 2021. *Meddling in the ballot box : the causes and effects of partisan electoral interventions*.

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

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

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.
URL: *https://www.journals.uchicago.edu/doi/10.1086/707308*

O'Rourke, Lindsey A. 2018. *Covert regime change : America's secret Cold War*.

Paine, Jack and Scott A Tyson. 2020. "Uses and abuses of formal models in political science.".

Pischedda, Costantino and Andrew Cheon. 2023. "Does plausible deniability work? Assessing the effectiveness of unclaimed coercive acts in the Ukraine war." *Contemporary Security Policy* 44:345–371.

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.

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." p. 237.

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

Ruachhaus, Robert W. 2009. "Principal-Agent Problems in Humanitarian Intervention: Moral Hazards, Adverse Selection, and the Commitment Dilemma." *International Studies Quarterly* 53(4):871–884.
URL: *http://www.jstor.org/stable/27735127*

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.

Seneter, 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.".

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

Taylor, Philip B., Jr. 1967. "Progress in Venezuela." *Current History* 53(315):270–274.
URL: *https://doi.org/10.1525/curh.1967.53.315.270*

Terman, Rochelle and Joshua Byun. 2022. "Punishment and Politicization in the International Human Rights Regime." *American Political Science Review* 116:385–402.

Todd, Chuck. 2018. "Republicans break from Trump on migrant family separation." *NBC* .

Wolford, Scott and Toby J. Rider. 2024. "Weak sovereignty and interstate war." *International Theory* 16:153–177.

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

6 Formal Appendix

Notation: We will reference the following notation throughout the proofs that follow:

- Let $r^{a_p,y,z}$ denote A's strategy as a function of a_p, y, z
- Let $q = r^{010}$, $s = r^{110}$, $t = r^{100}$, $v = r^{000}$
- Denote L's action $a = (a_p, a_c)$
- $k_c^\dagger := \hat{k}_c(q = 0, v = \bar{r}) = \alpha_c - \alpha_0 + \bar{r}((1-\lambda)(1-\alpha_c) - (1-\alpha_0))$
- $\lambda'' = \frac{(\alpha_c - \alpha_0)(1-\bar{r}) - \underline{k_c}}{\bar{r}(1-\alpha_c)}$
- $\hat{\lambda} = 1 - \underline{k_p}/\bar{r}$
- $\lambda' = 1 - \underline{k_p}/(\alpha_c \bar{r})$
- $\bar{\lambda} = (\alpha_c - \alpha_0 - \underline{k_c})/\bar{r}$

Restating the parameter restrictions from the main text:

Assumption 2 *Throughout the analysis, we assume:*

- $\bar{r} < 1$*: the leader values the policy outcome more than whatever reward/punishment the audience can impose on her.*
- $\alpha_0 < \min\{\underline{k_p}/\bar{r},\ \alpha_c(1-\lambda)\}$*: the likelihood of policy success due to exogenous factors is relatively small.*
- $\underline{k_p} \leq k_p^\dagger = \alpha_c(\alpha_p - \alpha_c + \underline{k_c} + \bar{r})$*: this ensures that the leader is willing to take public action in the favorable state, when she is not punished for doing so.*

6.1 L's best response – general case

We will first characterize the leader's best response as a function of the audience's strategy. Note that this best response function remains the same across the two versions of the model (where it characterizes the unscrupulous leader's behavior in the second version). The difference across the two models pertains only to the audience's payoff function.

Lemma 1 *The values $(\tilde{k}_p^1, \hat{k}_p, \tilde{k}_p^0, \hat{k}_c, \tilde{k}_c^1, \hat{k}_c^0, \ddot{k}_c^0, \tilde{k}_c^0)$, defined in the proof below, are functions of A's strategy $r = (q, s, t, v)$. For any A strategy r in which $r^{a_p,y,z=1} = 0$, L's best response is characterized as follows:*

If $k_p = \overline{k_p}$:

- *L plays (0,1) if $k_c < \hat{k}_c$, and (0,0) otherwise*

If $\omega = 1, k_p = \underline{k_p}$*:*

- *L never plays (1,1)*
- *if* $\tilde{k}_p^1 \geq k_p$*:*
  - *L plays (0,1) if* $k_c < \tilde{k}_c^1$*, and (1,0) otherwise*
- *if* $\tilde{k}_p^1 < k_p$*:*
  - *L plays (0,1) if* $k_c < \hat{k}_c$*, and (0,0) otherwise*

If $\omega = 0, k_p = \underline{k_p}$*:*

- *if* $k_p > \max\{\tilde{k}_p^0, \hat{k}_p\}$*:*
  - *L plays (0,1) if* $k_c < \hat{k}_c$*, and (0,0) otherwise*
- *if* $\tilde{k}_p^0 < k_p \leq \hat{k}_p$*:*
  - *L plays (1,1) if* $k_c < \hat{k}_c^0$*, and (0,0) otherwise*
- *if* $\hat{k}_p < k_p \leq \tilde{k}_p^0$*:*
  - *L plays (0,1) if* $k_c < \tilde{k}_c^0$*, and (1,0) otherwise*
- *if* $k_p \leq \min\{\hat{k}_p, \tilde{k}_p^0\}$*:*
  - *L plays (1,1) if* $k_c < \ddot{k}_c^0$*, and (1,0) otherwise*

Proof of Lemma 1: First, given that $Pr(y = 1|a_p\omega = 1) = Pr(y = 1|a_p\omega = 1, a_c = 1) = \alpha_p > Pr(y = 1|a_c = 1) = \alpha_c$, it is clear that L will never play $a = (1, 1)$ when $(\omega = 1, k_p = \underline{k_p})$, as this is strictly dominated by $a = (1, 0)$ (given that A punishes when covert action is revealed, as we show below).

When $\omega = 1$, $k_p = \underline{k_p}$:

$$EU_L(1, 0) = \alpha_p(1 + s) + (1 - \alpha_p)t - k_p$$

When $\omega = 0$, $k_p = \underline{k_p}$:

$$EU_L(1, 1) = \alpha_c(1 + (1 - \lambda)s) + (1 - \alpha_c)((1 - \lambda)t) - k_p - k_c$$
$$EU_L(1, 0) = \alpha_0(1 + s) + (1 - \alpha_0)t - k_p$$

In all states (including $k_p = \overline{k_p}$):

$$EU_L(0,1) = \alpha_c(1 + (1-\lambda)q) + (1-\alpha_c)(1-\lambda)v - k_c$$
$$EU_L(0,0) = \alpha_0(1+q) + (1-\alpha_0)v$$

We will proceed to characterize L's preferences between each of these options (conditional on ω, k_p) in terms of thresholds in either k_c or k_p, where the thresholds are functions of A's strategy $r = (q, s, t, v)$.

Threshold for (0,0) vs. (0,1):

$$EU_L(0,0) < EU_L(0,1) \iff k_c < \hat{k}_c, \quad \uparrow q, v$$
$$\hat{k}_c := \alpha_c - \alpha_0 + q[\alpha_c(1-\lambda) - \alpha_0] + v[(1-\alpha_c)(1-\lambda) - (1-\alpha_0)]$$

Thresholds when $\omega = 1$, $k_p = \underline{k_p}$:

$$EU_L(0,0) < EU_L(1,0) \iff k_p < \tilde{k}_p^1, \quad \uparrow s, t, \quad \downarrow q, v$$
$$\tilde{k}_p^1 := \alpha_p - \alpha_0 + \alpha_p s + (1-\alpha_p)t - \alpha_0 q - (1-\alpha_0)v$$

$$EU_L(1,0) < EU_L(0,1) \iff k_c < \tilde{k}_c^1, \quad \uparrow q, v, \quad \downarrow s, t$$
$$\tilde{k}_c^1 := \alpha_c - \alpha_p + k_p + \alpha_c(1-\lambda)q + (1-\alpha_c)(1-\lambda)v - \alpha_p s - (1-\alpha_p)t$$

Thresholds when $\omega = 0$, $k_p = \underline{k_p}$:

$$EU_L(0,0) < EU_L(1,0) \iff k_p < \tilde{k}_p^0, \quad \uparrow s, t, \quad \downarrow q, v$$
$$\tilde{k}_p^0 := \alpha_0(s-q) + (1-\alpha_0)(t-v)$$

$$EU_L(0,1) < EU_L(1,1) \iff k_p < \hat{k}_p, \quad \uparrow s, t, \quad \downarrow q, v$$
$$\hat{k}_p := (1-\lambda)[\alpha_c s + (1-\alpha_c)t] - (1-\lambda)[\alpha_c q + (1-\alpha_c)v]$$

$$EU_L(0,0) < EU_L(1,1) \iff k_c < \hat{k}_c^0, \quad \uparrow s, t, \quad \downarrow q, v$$
$$\hat{k}_c^0 := \alpha_c - \alpha_0 - k_p + (1-\lambda)[\alpha_c s + (1-\alpha_c)t] - \alpha_0 q - (1-\alpha_0)v$$

Figure 4: L's best response when $\omega = 0, k_p = \underline{k_p}$

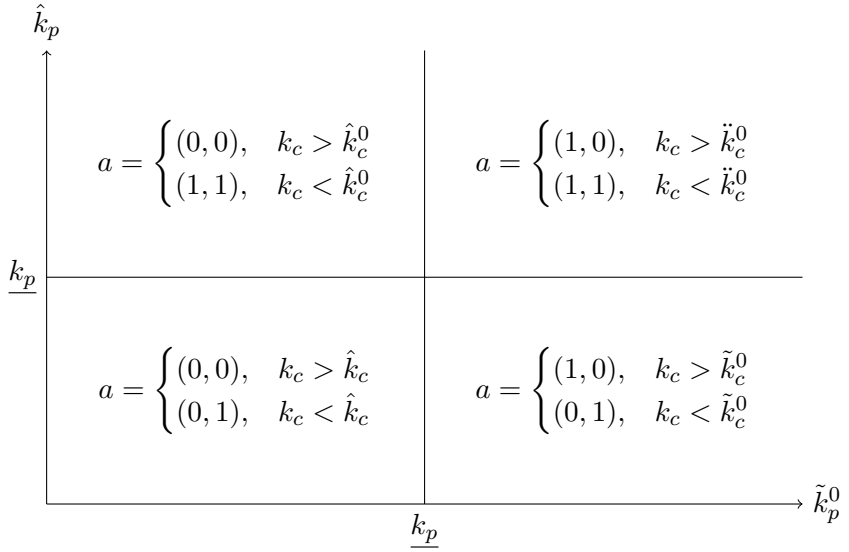
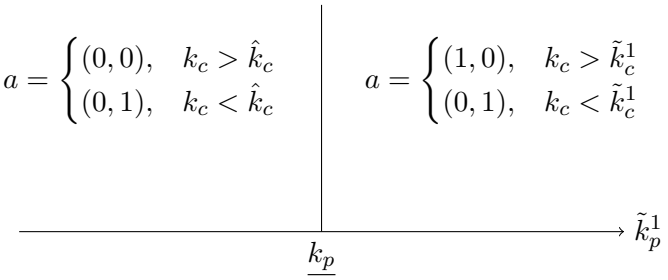


Figure 5: L's best response when $\omega = 1, k_p = \underline{k_p}$



$$EU_L(1,0) < EU_L(0,1) \iff k_c < \tilde{k}_c^0, \quad \uparrow q, v, \quad \downarrow s, t$$
$$\tilde{k}_c^0 := \alpha_c - \alpha_0 + (1-\lambda)[\alpha_c q + (1-\alpha_c)v] - \alpha_0 s - (1-\alpha_0)t + k_p$$

$$EU_L(1,0) < EU_L(1,1) \iff k_c < \ddot{k}_c^0, \quad \uparrow s, \quad \downarrow t$$
$$\ddot{k}_c^0 := \alpha_c - \alpha_0 + s[(1-\lambda)\alpha_c - \alpha_0] + t[(1-\lambda)(1-\alpha_c) - (1-\alpha_0)]$$

■

Figures 4 and 5 provide a visualization of the leader's best-response, as characterized in Lemma 1.

Remark

- $\tilde{k}_p^0 < \tilde{k}_p^1$
- If $\hat{k}_p \geq k_p$, then $\tilde{k}_c^1 < 0$.

- $\hat{k}_c - \tilde{k}_c^1 = \tilde{k}_p^1 - \underline{k_p}$
- $\hat{k}_c^0 - \hat{k}_c = \hat{k}_p - \underline{k_p}$
- $\tilde{k_c}^1 - \tilde{k}_c^0 = \hat{k}_p - \underline{k_p}$
- $\hat{k}_c - \tilde{k}_c^0 = \tilde{k}_p^0 - \underline{k_p}$

6.2 Analysis: Homogeneous Leader Types (Pure Moral Hazard)

Before proving Proposition 1 from the main text, it will be useful to establish the following lemmas:

Lemma 2 *Consider two A strategies $r' = (q', s', t', v')$ and $r'' = (q'', s'', t'', v'')$. We say that $r' < r''$ if $q' \leq q''$, $s' \leq s''$, $t' \leq t''$, and $v' \leq v''$, with at least one strict inequality. If (i) $r' < r''$; and (ii) L's best response to either one yields an equilibrium probability of public action $p \in (0, 1)$; then r' yields L a strictly higher payoff than does r''.*

This lemma tells us that in order to satisfy Assumption 1, we can first look for A strategies that minimize L's probability of taking covert action; and then for any r' and r'' among these strategies for which $r' < r''$, eliminate r''.

Lemma 3 *If there exists an equilibrium satisfying Assumption 1 in which $r^{a_p', y', z=1} > 0$ for some a_p', y'; then there also exists an equilibrium satisfying Assumption 1 with the same probability of covert action, and the same L payoff, in which $r^{a_p, y, z=1} = 0$.*

Proof of Lemma 3: If $r^{a_p', y', z=1} > 0$ and L takes covert action with positive probability when playing $a_p = a_p'$, then A can reduce the probability of covert action by setting $r^{a_p', y', z=1} = 0$ (which means that $r^{a_p', y', z=1} > 0$ was not part of an equilibrium satisfying Assumption 1). If $r^{a_p', y', z=1} > 0$ but L never takes covert action with positive probability when playing $a_p = a_p'$ (meaning $(a_p', y', z = 1)$ is off the equilibrium path of play), then A setting $r^{a_p', y', z=1} = 0$ does not violate Assumption 1. ■

Given Lemma 3, it will be without loss of generality to restrict attention to equilibria in which $r^{a_p', y', z=1} = 0$ for all a_p', y'.

Proposition 1 makes multiple claims. We will reproduce the proposition here, and break down each claim into its own proposition to prove separately.

Proposition 1 (Equilibria under pure moral hazard) In the pure moral hazard game, there exist thresholds λ' and λ'' of the transparency parameter λ such that:

- If $\lambda \geq \lambda''$, then in any equilibrium satisfying Assumption 1, the leader never takes covert action. (Proposition 3)
- If $\lambda \in (\lambda', \lambda'')$, then in any equilibrium satisfying Assumption 1, the leader takes covert action with positive probability but never uses a cover story. (Proposition 4)
- If $\lambda < \min\{\lambda', \lambda''\}$, there exists an equilibrium satisfying Assumption 1 in which the leader uses a cover story with positive probability. (Proposition 5)

Proposition 3 *Let* $\lambda'' := \frac{(\alpha_c - \alpha_0)(1-\bar{r}) - \underline{k_c}}{\bar{r}(1-\alpha_c)}$. *If* $\lambda \geq \lambda''$, *then in any equilibrium satisfying Assumption 1, L never plays covert action.*

Proof of Proposition 3: We will show that there exists an A strategy that induces an equilibrium probability of covert action of zero. Thus any equilibrium satisfying Assumption 1 must have zero probability of covert action.

Consider the following A strategy: $v = \bar{r}$, $q = 0$, and s, t that maximize $\alpha_c s + (1 - \alpha_c)t$ subject to

$$\alpha_c s + (1 - \alpha_c)t \leq \frac{\underline{k_p}}{1 - \lambda} + (1 - \alpha_c)\bar{r} \tag{7}$$

Under this strategy, we have the following:

- $\hat{k}_p \leq \underline{k_p}$ (which is simply a rearrangement of (7))
- $\tilde{k}_p^0 < \underline{k_p}$ (given $\alpha_0 \bar{r} < \underline{k_p}$, as per Assumption 2)
- $\hat{k}_c = k_c^\dagger \leq \underline{k_c}$, given $\lambda \geq \lambda''$
- $\tilde{k}_c^1 < 0$ and $\tilde{k}_p^1 > \underline{k_p}$ (as per Remark 6.1)

Thus we have $Pr(a_c = 1) = (1 - \tau_0 - \tau_1)F(\hat{k}_c) + \tau_0 F(\hat{k}_c) + \tau_1 F(\tilde{k}_c^1) = 0$.

Since there exists a strategy by A that eliminates the risk of covert action entirely, we know that in any optimal equilibrium for A, there is zero probability of covert action being taken. (To satisfy Assumption 1, A may increase s, t, and q to the highest possible values that maintain zero probability of covert action.) ■

Lemma 4 *If* $\lambda < \lambda''$, *then A can achieve a covert action probability of* $(1 - \tau_1)F(k_c^\dagger)$

Proof of Lemma 4: If A plays $q = 0, v = \bar{r}$, then $\tilde{k}_p^0 < \underline{k_p}$, and $\hat{k}_c = k_c^\dagger$. Then there are two cases to consider.

(1) If $\lambda > \lambda' = 1 - \underline{k_p}/(\alpha_c \bar{r})$:

– Consider an A strategy of $(s = t = v = \bar{r}, q = 0)$. Under this strategy, we have $\tilde{k}_p^0 \leq \underline{k_p}$, $\hat{k}_p < \underline{k_p}$, and $\tilde{k}_p^1 > \underline{k_p}$. Because $\underline{k_p} \leq k_p^\dagger$, we have that $\tilde{k}_c^0 \leq \underline{k_c}$.

– Thus $Pr(a_c = 1) = (1 - \tau_1)F(\hat{k}_c) + \tau_1 F(\tilde{k}_c^1)$, which reduces to $(1 - \tau_1)F(k_c^\dagger)$.

(2) If $\lambda \leq \lambda'$:

– Then $\hat{k}_p(s = t = v = \bar{r}, q = 0) \geq \underline{k_p}$, and there exist s', t' such that $\hat{k}_p(s = s', t = t', v = \bar{r}, q = 0) = \underline{k_p}$. For $r = (s', t', v = \bar{r}, q = 0)$, we have $\tilde{k}_p^0 \leq \underline{k_p}$, $\tilde{k}_c^1 < 0$, and $\tilde{k}_p^1 > \underline{k_p}$. So again we have $Pr(a_c = 1) = (1 - \tau_1)F(k_c^\dagger)$.

■

Lemma 5 *If $\lambda < \min\left\{\hat{\lambda}, \lambda''\right\}$, then in any equilibrium satisfying Assumption 1, we have $\hat{k}_p \leq \underline{k_p}$ and $\tilde{k}_p^0 \leq \underline{k_p}$.*

Proof of Lemma 5:

We will consider four cases: (1) $\hat{k}_p > \underline{k_p}$ and $\tilde{k}_p^0 > \underline{k_p}$; (2) $\tilde{k}_p^0 \leq \underline{k_p} < \hat{k}_p$; (3) $\hat{k}_p \leq \underline{k_p} < \tilde{k}_p^0$; and (4) $\hat{k}_p \leq \underline{k_p}$ and $\tilde{k}_p^0 \leq \underline{k_p}$. We will show that only (4) can be supported in an equilibrium satisfying Assumption 1.

(1) If $\hat{k}_p > \underline{k_p}$ and $\tilde{k}_p^0 > \underline{k_p}$:

- This implies $\tilde{k}_c^1 < 0$, so $\tilde{k}_p^1 > \underline{k_p}$ (as per Remark 6.1).
- Then $Pr(a_c) = \tau_0 F(\ddot{k}_c^0) + \tau_1(0) + (1 - \tau_0 - \tau_1)F(\hat{k}_c)$, where $\hat{k}_c$ is decreasing in v and $\ddot{k}_c^0$ is constant in v.
- Thus Assumption 1 requires that A increase v until either $\hat{k}_p \leq \underline{k_p}$ or $\tilde{k}_p^0 \leq \underline{k_p}$. So this case cannot be supported in equilibrium.

(2) If $\tilde{k}_p^0 \leq \underline{k_p} < \hat{k}_p$:

- This again implies $\tilde{k}_c^1 < 0$, so $\tilde{k}_p^1 > \underline{k_p}$ (as per Remark 6.1).
- Then $Pr(a_c) = \tau_0 F(\hat{k}_c^0) + \tau_1(0) + (1 - \tau_0 - \tau_1)F(\hat{k}_c)$, where $\hat{k}_c^0$ is strictly increasing in s, t and $\hat{k}_c$ is constant in s, t. (Note that $\hat{k}_p > \underline{k_p} \implies \hat{k}_c^0 > \hat{k}_c$, and $\lambda < \lambda'' \implies \hat{k}_c > \underline{k_c}$.)
- Thus Assumption 1 requires that A decrease s and/or t until $\hat{k}_p \leq \underline{k_p}$. (Note that $\hat{k}_p$ is increasing in s, t, and $\hat{k}_p(s \leq q, t \leq v) < 0 < \underline{k_p}$.) So this case cannot be supported in equilibrium.

(3) If $\hat{k}_p \leq \underline{k_p} < \tilde{k}_p^0$:

- Note that $\tilde{k}_p^0 < \tilde{k}_p^1$, so this condition means that $Pr(a_c) = (1-\tau_1-\tau_0)F(\hat{k}_c) + \tau_0 F(\tilde{k}_c^0) + \tau_1 F(\tilde{k}_c^1)$, which is strictly increasing in q, and weakly decreasing in s, t.
- If $\hat{k}_p(s=t=\bar{r}, q=0, v) < \underline{k_p}$ and $\tilde{k}_p^0(s=t=\bar{r}, q=0, v) > \underline{k_p}$ for some v:
  - Then Assumption 1 requires that any equilibrium with $\hat{k}_p \leq \underline{k_p} < \tilde{k}_p^0$ must feature $s=t=\bar{r}$, $q=0$.
  - Given $s=t=\bar{r}$, $q=0$, and $\tilde{k}_p^0 > \underline{k_p}$, it follows that $\tilde{k}_c^1 < \underline{k_c}$.
  - Thus $Pr(a_c) = (1-\tau_0-\tau_1)F(\hat{k}_c) + \tau_0 F(\tilde{k}_c^0) + \tau_1(0)$. Because $\hat{k}_c > \underline{k_c}$ and $\tilde{k}_c^0 > \underline{k_c}$, we know that $Pr(a_c)$ is monotonic in v.
  - If $Pr(a_c)$ is decreasing in v, then Assumption 1 requires that A increase v until $\tilde{k}_p^0 \leq \underline{k_p}$. Then the condition of $\underline{k_p} < \tilde{k}_p^0$ cannot e supported in equilibrium.
  - Otherwise, if $Pr(a_c)$ is increasing in v, Assumption 1 requires that A decrease v until $\hat{k}_p = \underline{k_p}$ (which we know can be satisfied, given that $\lambda < \hat{\lambda} \iff \hat{k}_p(s=t=\bar{r}, q=v=0) > \underline{k_p}$).
- Thus any equilibrium with $\hat{k}_p \leq \underline{k_p} < \tilde{k}_p^0$ must satisfy $\hat{k}_p = \underline{k_p}$.
  - Given $\hat{k}_p = \underline{k_p} < \tilde{k}_p^0$, we know that $Pr(a_c) = (1-\tau_0-\tau_1)F(\hat{k}_c) + \tau_0 F(\tilde{k}_c^0) + \tau_1(0)$.
  - Further, $\hat{k}_p = \underline{k_p}$ implies that $\tilde{k}_c^0 = \alpha_c - \alpha_0 + (1-\lambda)(\alpha_c s + (1-\alpha_c)t) - \alpha_0 s - (1-\alpha_0)t$. So $\tilde{k}_c^0 \geq k_c^\dagger$, with equality only for $s=0, t=\bar{r}$; but in order to support $\hat{k}_p = \underline{k_p}$ with $s=0$, we would require $v < \bar{r}$, which means $\hat{k}_c > k_c^\dagger$. This implies a strictly higher probability of covert action than $(1-\tau_1)F(k_c^\dagger)$, which we know can be achieved as per Lemma 4.

Altogether, given $\lambda < \min\left\{\hat{\lambda}, \lambda''\right\}$, any equilibrium satisfying Assumption 1 must be characterized by $\hat{k}_p \leq \underline{k_p}$ and $\tilde{k}_p^0 \leq \underline{k_p}$.

■

Lemma 6 *If $\lambda < \min\left\{\hat{\lambda}, \lambda''\right\}$, then in any equilibrium satisfying Assumption 1, A will play $v = \bar{r}, q = 0$.*

Proof of Lemma 6: Following Lemma 5, we know that in equilibrium, $\hat{k}_p \leq \underline{k_p}$ and $\tilde{k}_p^0 \leq \underline{k_p}$. Further, it must be the case that $\tilde{k}_p^1 \geq \underline{k_p}$; if not, A would have a strict incentive to set $v = \bar{r}$, and then increase s and/or t until $\tilde{k}_p^1 = \underline{k_p}$. So $Pr(a_c) = (1-\tau_1)F(\hat{k}_c) + \tau_1 F(\tilde{k}_c^1)$. Observe that $F(\hat{k}_c)$ is uniquely minimized by $(v=\bar{r}, q=0)$; and given $(v=\bar{r}, q=0)$, $F(\tilde{k}_c^1)$ can be brought to zero for sufficiently high s and/or t. Therefore, minimizing $Pr(a_c)$ requires $(v=\bar{r}, q=0)$. ■

Proposition 4 *If $\lambda' < \lambda < \lambda''$, then in any equilibrium satisfying Assumption 1, L plays covert action with positive probability, but never uses a cover story.*

Proof of Proposition 4: We'll prove the proposition in two cases: first, $\hat{\lambda} < \lambda < \lambda''$; and second, $\lambda' < \lambda \leq \hat{\lambda}$.

(1) If $\hat{\lambda} < \lambda < \lambda''$:

- Note that $\hat{\lambda} < \lambda < \lambda''$ rearranges to: $\underline{k_p} > \bar{r}(1-\lambda)$, and $\hat{k}_c > \underline{k_c}$ for all q, v. In the $k_p = \overline{k_p}$ state, L takes covert action iff $k_c < \hat{k}_c$, which occurs with positive probability.
- From Lemma 1, we know that the use of a cover story—that is, L playing (1,1) when $\omega = 0, k_p = \underline{k_p}$—requires $\hat{k}_p \geq \underline{k_p}$. But the highest possible value of $\hat{k}_p$ is $\bar{r}(1-\lambda)$, which is less than $\underline{k_p}$, under the conditions of the proposition.

(2) If $\lambda' < \lambda \leq \hat{\lambda}$:

- Note that this rearranges to

$$\hat{k}_p(s = t = v = \bar{r}, q = 0) = \bar{r}(\alpha_c(1-\lambda)) < \underline{k_p} \leq \bar{r}(1-\lambda) = \hat{k}_p(s = t = \bar{r}, q = v = 0)$$

- Following Lemma 5, we know that in equilibrium, $q = 0$ and $v = \bar{r}$. Any such strategy implies $\hat{k}_p < \underline{k_p}$ (given $\lambda > \lambda'$), meaning L will never use a cover story.

■

Proposition 5 *If $\lambda < \min\{\lambda', \lambda''\}$, then in any equilibrium satisfying Assumption 1:*

- *A plays $q = 0$, $v = \bar{r}$, and $s, t : \alpha_c s + (1-\alpha_c)t = \frac{k_p}{1-\lambda} + (1-\alpha_c)\bar{r}$*
- *L plays covert action with probability*

$$Pr(a_c = 1) = (1-\tau_1)F(k_c^\dagger), \text{ where } k_c^\dagger = \alpha_c - \alpha_0 + \bar{r}((1-\alpha_c)(1-\lambda) - (1-\alpha_0)$$

- *L can play a cover story with any probability.*

Proof of Proposition 5: In any equilibrium satisfying Assumption 1:

- Lemma 5 showed that it must be the case that $\hat{k}_p \leq \underline{k_p}$ and $\tilde{k}_p^0 \leq \underline{k_p}$.
- Lemma 6 showed that it must be the case that $v = \bar{r}$ and $q = 0$.

This means that

$$Pr(a_c) = (1-\tau_1)F(k_c^\dagger) + \tau_1 Pr(a_c|\omega = 1, \underline{k_p})$$

Left to be determined in A's strategy are the values of s and t. We can see that $Pr(a_c|\omega = 1, \underline{k_p})$ is weakly decreasing in s and t. So any equilibrium satisfying both Assumption 1, and the condition

that $\hat{k}_p \leq \underline{k_p}$, requires that s and t satisfy $\alpha_c s + (1-\alpha_c)t = \frac{\underline{k_p}}{1-\lambda} + (1-\alpha_c)\bar{r}$ (which gives $\hat{k}_p = \underline{k_p}$): if either s or t were higher, then $\hat{k}_p \leq \underline{k_p}$ would be violated, and if s or t were lower, then payoff-optimality for L would be violated. Observe that this strategy implies $\tilde{k}_p^1 > \underline{k_p}$ and $\tilde{k}_c^1 < \underline{k_c}$, so $Pr(a_c|\omega = 1, \underline{k_p}) = 0$. ■

6.3 Analysis: Heterogeneous Leader Types (Adverse Selection)

Lemma 7 *In the heterogeneous leader model:*

- *The unscrupulous leader's best response is as characterized in Lemma 1.*
- *The audience's strategy must satisfy*

$$r = \begin{cases} 0, & \mu^{a_p,y,z} < \bar{\mu} \\ \bar{r}, & \mu^{a_p,y,z} > \bar{\mu} \\ r \in [0, \bar{r}], & \mu^{a_p,y,z} = \bar{\mu} \end{cases} \tag{8}$$

  where $\mu^{a_p,y,z}$ denotes the audience's posterior belief that the leader is a scrupulous type, having observed a_p, y, z.

- *If $a_c = 1$ occurs with positive probability on the equilibrium path of play, then $\mu^{a_p,y,z=1} = 0$ for all a_p, y.*

We will impose the following assumption throughout the analysis:

Assumption 3 *If $z = 1$ is off the equilibrium path of play, the audience assigns probability zero to the leader being scrupulous upon observing $z = 1$.*

Note that because the scrupulous leader is defined as a behavioral type (and because the probabilities of policy success $(\alpha_0, \alpha_p) \in (0, 1)$), every information set with $z = 0$ occurs on the equilibrium path of play, so the audience's beliefs in those information sets can be determined by Bayes' Rule.

Proposition 2 makes multiple claims. We will reproduce the proposition here, and break down each claim into its own proposition to prove separately.

Proposition 2 (Equilibria under adverse selection) In the adverse selection game, there exist thresholds λ' and $\bar{\lambda}$ of the transparency parameter λ such that:

- If $\lambda \geq \bar{\lambda}$, there exists an equilibrium with zero probability of covert action. (Proposition 6)
- If $\lambda' < \lambda < \bar{\lambda}$, then in any equilibrium, the leader takes covert action with positive probability, but never uses a cover story. (Proposition 7)

- If $\lambda < \min\{\lambda', \lambda''\}$, and if $(\pi - \bar{\mu})$ is not too large, then (Proposition 8):
  - There exists an equilibrium in which L plays a cover story with positive probability, and this equilibrium yields the lowest possible probability of covert action.
  - Any equilibrium featuring the lowest possible probability of covert action requires that a cover story be played with strictly positive probability. Further, note that in such an equilibrium, if $(1 - \tau_0 - \tau_1)$ is not too large, then L must play a cover story as part of a completely mixed strategy. That is: conditional on taking covert action in the $(\omega = 1, k_p = \underline{k_p})$ state, L must place positive probability on both $a = (0, 1)$ and $a = (1, 1)$.

Proposition 6 *If $\lambda \geq \bar{\lambda} = (\alpha_c - \alpha_0 - \underline{k_c})/\bar{r}$, there exists an equilibrium with zero probability of covert action. If $\lambda < \bar{\lambda}$, no such equilibrium exists.*

Proof of Proposition 6:

Note that $\lambda \geq \bar{\lambda}$ rearranges to $\hat{k}_c(q = v = \bar{r}) \leq \underline{k_c}$.

First, if $\lambda \geq \bar{\lambda}$: consider the A strategy $(s = t = q = v = \bar{r})$, and the L strategy

$$a_c = 0, \quad a_p = \begin{cases} 1, & \omega = 1, k_p = \underline{k_p} \\ 0 & otw \end{cases}$$

This L strategy implies $\mu^{a_p, y, z=0} = \pi > \bar{\mu}$ for all a_p, y, which renders A's strategy a best response with consistent beliefs. A's strategy in turn yields $\tilde{k}_p^0 < \underline{k_p}$, $\hat{k}_p < \underline{k_p}$, $\tilde{k}_p^1 > \underline{k_p}$, $\tilde{k}_c^1 < 0$, and $\hat{k}_c = \alpha_c - \alpha_0 - \bar{r}\lambda \leq \underline{k_c}$. Thus there is zero probability of covert action if $\lambda \geq \bar{\lambda}$.

Next, we will show that if $\lambda < \bar{\lambda}$, then there is no equilibrium with zero probability of covert action. In the $k_p = \overline{k_p}$ state, the probability of covert action is $F(\hat{k}_c)$, which we will show must be strictly positive given $\lambda < \bar{\lambda}$.

Suppose that $\lambda < \bar{\lambda}$ and there is an equilibrium with no covert action. This implies that

$$\mu^{000} = \mu^{010} = \frac{\pi}{\pi + (1 - \pi)\left[\frac{\tau}{1-\tau}(1 - \sigma^1) + (1 - \sigma^0)\right]}$$

where $\sigma^1 = Pr(a_p = 1 | \omega = 1 \text{ and } k_p = \underline{k_p})$, and $\sigma^0 = Pr(a_p = 1 | \omega = 0 \text{ or } k_p = \overline{k_p})$, in the unscrupulous leader's strategy. Note that in the scrupulous leader's strategy, these probabilities are simply 1 and 0, respectively.

Thus unless $\mu^{000} = \mu^{010} = \bar{\mu}$, it must be the case that $q = v$. If $q = v$ then $\hat{k}_c = \alpha - c - \alpha_0 - \lambda q$, which is $> \underline{k_c}$ given that $\lambda > \bar{\lambda}$. Thus an equilibrium with zero covert action requires $\mu^{000} = \mu^{010} = \bar{\mu}$ and $v > q$.

$\mu^{0y0} = \bar{\mu}$ implies that $\frac{\tau_1}{1-\tau_1}(1-\sigma^1) + (1-\sigma^0) > 1$, which rearranges to $\tau_1(1-\sigma^1) > (1-\tau_1)\sigma^0$. This means that the unscrupulous leader plays $a_p = 1$ with probability strictly less than one when $(\omega = 1, k_p = \underline{k_p})$, that is, $\sigma^1 < 1$.

$\sigma^1 < 1$ implies $\tilde{k}_p^1 \leq \underline{k_p}$, which means $\tilde{k}_p^0 < \underline{k_p}$, which in turn implies $\sigma^0 = 0$. It follows that both $\mu^{110} > \bar{\mu}$ and $\mu^{100} > \bar{\mu}$, so $s = t = \bar{r}$. But this means $\tilde{k}_p^1 > \underline{k_p}$, contradicting $\tilde{k}_p^1 \leq \underline{k_p}$ as stated above.

This contradiction followed from the supposition that $\lambda < \bar{\lambda}$ and there is an equilibrium with no covert action. Thus if $\lambda < \bar{\lambda}$, there cannot be an equilibrium with zero probability of covert action. ■

Remark Observe that $\bar{\lambda} > \lambda'$: the audience can induce zero covert action under a wider range of conditions in the homogeneous-types model, as compared to the heterogeneous-types model.

Proposition 7 *If $\lambda > \lambda' = 1 - \underline{k_p}/(\bar{r}\alpha_c)$, then there does not exist an equilibrium in which the leader uses a cover story.*

Proof of Proposition 7: $\lambda > \lambda'$ rearranges to $\underline{k_p} > \hat{k}_p(s = t = v = \bar{r}, q = 0)$. The use of a cover story requires $\hat{k}_p \geq \underline{k_p}$. The only way to increase $\hat{k}_p$ above $\hat{k}_p(s = t = v = \bar{r}, q = 0)$ is to lower v (given that $\hat{k}_p$ is increasing in s, t and decreasing in q).

However, $\hat{k}_p \geq \underline{k_p} \implies \tilde{k}_c^1 < 0 < \hat{k}_c \implies \tilde{k}_p^1 > \underline{k_p}$. This means that when $(\omega = 1, k_p = \underline{k_p})$, L always plays $(1, 0)$.

For shorthand, let $\bar{\omega}$ denote the state that either $\omega = 0$ or $k_p = \overline{k_p}$, and let $\underline{\omega}$ denote the state that $\omega = 1$ and $k_p = \underline{k_p}$, so $Pr(\bar{\omega}) = 1 - Pr(\underline{\omega})$. Also let $h = (0, 0, 0)$ denote the information set $(a_p = 0, y = 0, z = 0)$.

Then $\mu^{000} = \frac{\pi(1-\tau_1)(1-\alpha_0)}{\pi(1-\tau_1)(1-\alpha_0)+(1-\pi)\left[(1-\tau_1)Pr(h=(0,0,0)|\bar{\omega},\theta=0)+\tau_1 Pr(h=(0,0,0)|\underline{\omega},\theta=0)\right]}$.

But we know that:

- $Pr(h = (0, 0, 0)|\bar{\omega}, \theta = 0) \leq 1 - \alpha_0 = Pr(h = (0, 0, 0)|\bar{\omega}, \theta = 1)$, and
- $Pr(h = (0, 0, 0)|\underline{\omega}, \theta = 0) = 0$

So $\mu^{000} \geq \pi$. Which means that $v = \bar{r}$, and thus $\hat{k}_p < \underline{k_p}$, and we cannot have a cover story.

■

Lemma 8 *Suppose the unscrupulous leader plays a strategy characterized as follows:*

- *When $\omega = 1, k_p = \underline{k_p}$: L plays $a = (1, 0)$*
- *When $\omega = 0$ or $k_p = \underline{k_p}$: L plays $a = (0, 0)$ with total probability σ_{00}, and likewise σ_{01} and σ_{11}, with $\sigma_{00} + \sigma_{01} + \sigma_{11} = 1$*

Then, according to Bayes' Rule, the audience's beliefs $\mu^{a_p,y,z}$ satisfy:

$$\mu^{1,1,0} = \frac{\pi\tau_1\alpha_p}{\pi\tau_1\alpha_p + (1-\pi)[\tau_1\alpha_p + (1-\tau_1)\sigma_{11}(1-\lambda)\alpha_c]}$$
$$\mu^{1,0,0} = \frac{\pi\tau_1(1-\alpha_p)}{\pi\tau_1(1-\alpha_p) + (1-\pi)[\tau_1(1-\alpha_p) + (1-\tau_1)\sigma_{11}(1-\lambda)(1-\alpha_c)]}$$
$$\mu^{0,1,0} = \frac{\pi(1-\tau_1)\alpha_0}{\pi(1-\tau_1)\alpha_0 + (1-\pi)(1-\tau_1)[\sigma_{00}\alpha_0 + \sigma_{01}(1-\lambda)\alpha_c]}$$
$$\mu^{0,0,0} = \frac{\pi(1-\tau_1)(1-\alpha_0)}{\pi(1-\tau_1)(1-\alpha_0) + (1-\pi)(1-\tau_1)[\sigma_{00}(1-\alpha_0) + \sigma_{01}(1-\lambda)(1-\alpha_c)]}$$

Observe that $\mu^{0,0,0} \geq \pi$, and that $\mu^{1,1,0} > \mu^{1,0,0} \iff \alpha_p > \alpha_c$.

Proposition 8 *If $\lambda < \min\{\lambda', \lambda''\}$, and if $(\pi - \bar{\mu})$ is not too large, then:*

- *There exists an equilibrium in which L plays a cover story with positive probability, and this equilibrium yields the lowest possible probability of covert action.*
- *Any equilibrium featuring the lowest possible probability of covert action requires that a cover story be played with strictly positive probability.*
  - *In such an equilibrium, if $(1 - \tau_0 - \tau_1)$ is not too large, then L must play a cover story as part of a completely mixed strategy.*
    - *(That is: conditional on taking covert action in the $(\omega = 1, k_p = \underline{k_p})$ state, L must place positive probability on both $a = (0,1)$ and $a = (1,1)$.)*

Proof of Proposition 8: For the first point, we will show that an equilibrium meeting the conditions of the equilibrium characterized in Proposition 5—specifically, $q = 0, v = \bar{r}$, and $s, t : \alpha_c s + (1 - \alpha_c)t = \frac{\underline{k_p}}{1-\lambda} + (1-\alpha_c)\bar{r}$—can be supported under the heterogeneous leader type setting.

Proposition 5 showed that, when the audience's strategy was not pinned down by sequential rationality, any equilibrium that yields the lowest possible probability of covert action must satisfy these conditions. The equilibrium in this heterogeneous leader setting can simply be thought of as a refinement on the previously derived equilibrium, in which the audience's strategy must satisfy (8).

Consider the A strategy $s = v = \bar{r}, q = 0, t = \hat{t} \in (0, \bar{r})$, where $t = \hat{t}$, with[29]

$$\hat{t} = \frac{\underline{k_p}}{(1-\lambda)(1-\alpha_c)} + \bar{r}\left(\frac{1-2\alpha_c}{1-\alpha_c}\right), \quad \text{which gives} \quad \hat{k}_p(s = v = \bar{r}, q = 0, t = \hat{t}) = \underline{k_p}$$

L's best response, as per Lemma 1, is:

[29] This proof focuses on the case of $\underline{k_p} \geq \bar{r}(1-\lambda)(2\alpha_c - 1)$. If this condition is not satisfied, then a similar equilibrium with $v = \bar{r}, q = t = 0, s = \hat{s} \in (0, \bar{r})$ can be supported, with a cover story played with positive probability, and yielding the same probability of covert action.

- when $\omega = 1, k_p = \underline{k_p}$: $\quad a_p = 1, a_c = 0$
- when: $\omega = 0$: $\quad a_c = 1$ with probability $F(k_c^\dagger)$, and $a_c = 0$ with complementary probability
- when $\omega = 0$ and $k_p = \underline{k_p}$: when L plays $a_c = 1$, she can play $a_p = 1$ with any probability (since $\hat{k}_p = \underline{k_p}$, making L indifferent)

A's strategy requires $\mu^{100} = \bar{\mu}$, which rearranges to

$$\psi = \frac{\pi(1-\bar{\mu})}{(1-\pi)\bar{\mu}} = 1 + \left(\frac{1-\tau_1}{\tau_1}\right)(1-\lambda)\left(\frac{1-\alpha_c}{1-\alpha_p}\right)\sigma_{11} \tag{9}$$

where $\sigma_{11} = Pr(a = (1,1)|\theta = 0, \omega = 0)$ denotes the probability of the unscrupulous leader playing a cover story given $\omega = 0$.

To verify that this strategy profile is an equilibrium, we need to show:

- $\mu^{110} \geq \bar{\mu}$, which we can see is satisfied because $\mu^{110} > \mu^{100} = \bar{\mu}$
- $\mu^{000} \geq \bar{\mu}$, which is satisfied by the fact that the unscrupulous leader takes both covert and public action with greater probability than does the scrupulous leader, giving $\mu^{000} \geq \pi$
- $\mu^{010} = \frac{\pi}{\pi + (1-\pi)\left[\sigma_{00} + \sigma_{01}\frac{\alpha_c(1-\lambda)}{\alpha_0}\right]} \leq \bar{\mu}$, where σ_{00} and σ_{01} denote the probabilities with which the unscrupulous L plays $a = (0,0)$ and $a = (0,1)$, respectively, given $\omega = 0$. Note that $\mu^{010} \leq \bar{\mu}$ is satisfied for sufficiently large σ_{01}, or conversely, for sufficiently small σ_{11} (since $\sigma_{01} + \sigma_{11}$ must equal $F(k_c^\dagger)$)

If $(\pi - \bar{\mu})$ is sufficiently small (meaning ψ is sufficiently close to 1), then (9) is satisfied for a σ_{11} value small enough to satisfy $\mu^{010} \leq \bar{\mu}$.

For the second point, we will show that in the equilibrium described above, the probability of a cover story must be strictly positive.

Obseve that $\lambda < \lambda'$ rearranges to $\underline{k_p} < \hat{k}_p(s = t = v = \bar{r}, q = 0)$. Lemma 5 showed that the minimum-a_c equilibrium is characterized by $\hat{k}_p \leq \underline{k_p}$ and $\tilde{k}_p^0 \leq \underline{k_p}$. The condition $\hat{k}_p \leq \underline{k_p}$ requires that either $q > 0$, or at least one of $s < \bar{r}$ or $t < \bar{r}$.

Suppose L never plays a cover story. Since covert action is played with positive probability, we know $\mu^{010} < \pi$. If $(\pi - \bar{\mu})$ is sufficiently small, then $\mu^{010} < \bar{\mu}$ and thus $q = 0$. So we must have $s < \bar{r}$ or $t < \bar{r}$. But if L never plays a cover story then $\mu^{1y0} \geq \pi > \bar{\mu}$, so $s = t = \bar{r}$ and $q = 0$, contradicting the condition stated above. Thus it must be the case that L plays a cover story with positive probability.

For the third point, we will show that in the equilibrium described above, if $(1 - \tau_0 - \tau_1)$ is sufficiently small, then the probability of a cover story must be strictly strictly less than 1.

Suppose that any time L takes covert action in the $(\omega = 0, k_p = \underline{k_p})$ state, she plays a cover story; that is, $Pr(a = (0,1)|\omega = 0, k_p = \underline{k_p}) = 0$. Then A's belief upon observing $(a_p = 0, y = 1, z = 0)$ is

$$\mu^{010} = \frac{\pi}{\pi + (1-\pi)(\sigma_{00} + \sigma_{01}(1-\lambda)\frac{(1-\alpha_c)}{(1-\alpha_0)})}$$

where $\sigma_{01} = Pr(a = (0,1)|\omega = 0)$, which equals $(1-\tau_0-\tau_1)Pr(a = (0,1)|\omega = 0, k_p = \overline{k_p})$ under the equilibrium proposed. If $(1-\tau_0-\tau_1)$ is sufficiently small, then $\mu^{010} > \bar{\mu}$, and A must play $q = 0$. This would imply $\hat{k}_p < \underline{k_p}$, and thus L would have a strict incentive to deviate to never playing a cover story. Therefore it must be the case that when L takes covert action, she plays a cover story with probability less than 1.

■

Proof of Corollary 2: For the first point:

$$Pr(\theta = 1|a_p = 1) = \frac{\pi}{\pi + (1-\pi)(1 + \frac{(1-\tau_1)}{\tau_1}\sigma_{11})}$$
$$Pr(\theta = 1|a_p = 0) = \frac{\pi}{\pi + (1-\pi)(1-\sigma_{11})}$$

Clearly $Pr(\theta = 1|a_p = 1) < Pr(\theta = 1|a_p = 0)$ whenever $\sigma_{11} > 0$.

For the second point:

$$Pr(a_c = 1|a_p = 1) = \frac{Pr(a_c = 1 \ \& \ a_p = 1)}{Pr(a_p = 1)} = \frac{(1-\pi)(1-\tau_1)\sigma_{11}}{\pi\tau_1 + (1-\pi)(\tau_1 + (1-\tau_1)\sigma_{11})}$$
$$Pr(a_c = 1|a_p = 0) = \frac{Pr(a_c = 1 \ \& \ a_p = 0)}{Pr(a_p = 0)} = \frac{(1-\pi)(1-\tau_1)(1-\sigma_{11})}{\pi(1-\tau_1) + (1-\pi)(1-\tau_1)(1-\sigma_{11})}$$

The condition that $Pr(a_c = 1|a_p = 1) > Pr(a_c = 1|a_p = 0)$ rearranges to

$$\sigma_{11}[\pi(1-\tau_1) + (1-\pi)(1-\tau_1)(1-\sigma_{11})] > (1-\sigma_{11})[\pi\tau_1 + (1-\pi)(\tau_1 + (1-\tau_1)\sigma_{11})]$$

which then simplifies to

$$\sigma_{11} > \frac{\tau_1}{\pi + (1-\pi)\tau_1}$$

■

6.4 Extension: public action raises the risk of direct exposure

Proof of Corollary 3: First observe that for $\delta > 0$, the condition for L to be indifferent between (1,1) and (0,1) in the $\omega = 0, k_p = \underline{k_p}$ state is

$$\underline{k_p} = \hat{k}_p = (1 - \lambda - \delta)(\alpha_c s + (1 - \alpha_c)t) - (1 - \lambda)(\alpha_c q + (1 - \alpha_c)v)$$

We will show that, $\lambda \leq \tilde{\lambda}$ and $\delta > 0$, there exists an equilibrium with the following properties:

- In the $k_p = \overline{k_p}$ state, and in the $(\omega = 1, k_p = \underline{k_p})$ state, L's strategy remains unchanged (relative to the equilibrium characterized in Proposition 8)
- A plays $s = v = \bar{r},\ q = 0,\ t = \tilde{t} :=$, which implies $\hat{k}_p = \underline{k_p}$
- L plays a cover story with positive probability, which is increasing in δ over a non-empty range of δ

■

6.5 Extension: combined effectiveness of public and covert action

Proof of Corollary 4: Suppose we have an equilibrium where A plays the same strategy as in the adverse selection, low-λ case, that is: $s = v = \bar{r},\ q = 0,\ t = \hat{t} = \frac{k_p}{(1-\lambda)(1-\alpha_c)} + \bar{r}\left(\frac{1-2\alpha_c}{1-\alpha_c}\right)$. Then L's strategy in the $\omega = 0$ states remains the same. In the $\omega = 1, k_p = \underline{k_p}$ state, we still have $\tilde{k}_p^1 > \underline{k_p}$, so L must choose between (1,0) vs. (1,1). Her payoffs from each action are:

$$E[U_L(1,0)|\omega = 1] = \alpha_p(1 + s) + (1 - \alpha_p)t - \underline{k_p}$$
$$E[U_L(1,1)|\omega = 1] = (\alpha_p + \varepsilon)(1 + (1 - \lambda)s) + (1 - \alpha_p - \varepsilon)(1 - \lambda)t - \underline{k_p} - k_c$$

Thus L plays (1,1) iff

$$k_c < -\lambda(\alpha_p s + (1 - \alpha_p)t) + \varepsilon(1 + (1 - \lambda)(s - t))$$

So first, if

$$k_c^* := -\lambda(\alpha_p \bar{r} + (1 - \alpha_p)\hat{t}) + \varepsilon(1 + (1 - \lambda)(\bar{r} - \hat{t})) \leq \underline{k_c}$$

then L will never play (1,1) and we have the same path-of-play behavior as in the $\varepsilon = 0$ case. This condition requires $\varepsilon < \underline{\varepsilon}$ for some threshold value $\underline{\varepsilon}$.

If $\underline{\varepsilon} < \varepsilon < \overline{\varepsilon}$ for some other threshold $\overline{\varepsilon}$, then we can still support A playing the specified strategy. This will require:

- $\mu^{100} = \bar{\mu}$, which rearranges to

$$\psi = 1 - \underline{\sigma}_{11}\left[1 - (1 - \lambda)(1 - \frac{\varepsilon}{1 - \alpha_p})\right] + \left(\frac{1 - \tau_1}{\tau_1}\right)\bar{\sigma}_{11}(1 - \lambda)\left(\frac{1 - \alpha_c}{1 - \alpha_p}\right) \tag{10}$$

where $\underline{\sigma}_{11} = Pr(a = (1,1)|\theta = 0, \omega = \underline{\omega})$ and $\bar{\sigma}_{11} = Pr(a = (1,1)|\theta = 0, \omega = \bar{\omega})$, with $\underline{\omega}$ denoting the state ($\omega = 1$ and $k_p = \underline{k_p}$), and $\bar{\omega}$ denoting the state ($\omega = 0$ or $k_p = \overline{k_p}$).

- $\mu^{100} \leq \mu^{110}$, which rearranges to

$$\underline{\sigma}_{11}\varepsilon \leq \left(\frac{1-\tau_1}{\tau_1}\right)\bar{\sigma}_{11}(\alpha_p - \alpha_c) \tag{11}$$

- Isolating $\underline{\sigma}_{11}$ from (10), plugging in to (11) and rearranging, we have

$$\underline{\sigma}_{11}\left[-\lambda + \frac{\varepsilon(1-\lambda)}{\alpha_p - \alpha_c}\right] \leq \psi - 1 \tag{12}$$

To recap, the equilibrium satisfying the conditions of the corollary exists if (12) is satisfied.

Observe that $\underline{\sigma}_{11} = F(k_c^*)$ is increasing in ε. Thus the lefthand side of (12) is increasing in ε, and is satisfied for sufficiently small ε; that is, for $\varepsilon \leq \bar{\varepsilon}$, with $\bar{\varepsilon} > 0$.

This proves equilibrium existence. Finally, we can see that given the constraint that $\bar{\mu} = \mu^{100}$, which rearranges to

$$\psi - 1 + \underline{\sigma}_{11}\left[\lambda + (1-\lambda)\frac{\varepsilon}{1-\alpha_p})\right] = \left(\frac{1-\tau_1}{\tau_1}\right)\bar{\sigma}_{11}(1-\lambda)\left(\frac{1-\alpha_c}{1-\alpha_p}\right)$$

it must be the case that $\bar{\sigma}_{11}$ is increasing in ε.

■