Cover Stories

Michael F. Joseph
UCSD
mfjoseph@ucsd.edu

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

July 7, 2025

Abstract

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

Word count: 16,691

In March 1960, the CIA began organizing Cuban exiles to oust Fidel Castro. Eisenhower demanded that the CIA take extraordinary precautions to avoid direct evidence of 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 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 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, the New York Times began reporting on speculations that the CIA could be training and equipping an invasion force.[1]

Why would Eisenhower choose to attract suspicion, while implementing a deeply controversial policy he wanted to keep secret? The conventional wisdom dictates that he would not. A substantial body of research argues that governments maintain plausible deniability by avoiding *direct* evidence of their involvement in secret policies (Smith, 2019; Spaniel and Poznansky, 2018; Joseph and Poznansky, 2018; Carnegie, 2021; Yoder and Spaniel, 2022), and that overt actions can attract scrutiny and thus raise the risk that secret policies get exposed (Carson, 2018; Poznansky, 2020; Colaresi, 2012).

While important, the risk of direct exposure is not the only factor that governments consider. We uncover a countervailing incentive whereby governments, counterintuitively, pursue overt actions in order 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.

[1] Times (1961); Brewer (1961*b*)

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

We divide our analysis into four sections that collectively demonstrate the breadth and depth of what cover stories can contribute to our understanding of a wide range political phenomena. First, we develop a formal model that demonstrates how cover stories can resolve a common strategic problem studied by scholars of secrecy and international security (Spaniel and Poznansky, 2018; Canfil, 2022; Smith, 2019; Colaresi, 2012; Carnegie, 2021; Yoder and Spaniel, 2022; Kurizaki, 2007; Bils and Smith, 2025). In the model, a government can achieve a policy objective through two different, independently chosen means: a public action that an audience directly observes; and a secret action that is only observed with some probability. The government holds private information about the efficacy of each policy lever. The audience finds the secret policies to be more objectionable, and wants to prevent the government from using them. The government's challenge is to achieve its policy objectives, while maintaining plausible deniability for any actions that the audience disapproves of.

The model reveals that the government's optimal strategy for maintaining plausible deniability depends critically on the level of *transparency*, or the probability that direct evidence of covert action would be revealed. Under high transparency, the government is unlikely to use covert action; and when a successful policy outcome is achieved, the audience infers that the outcome likely came about without any unobserved intervention by the government. When transparency is low, however, the audience is no longer willing to give the government the "benefit of the doubt". Rather, if they observe a successful outcome despite public inaction, they will infer that covert action was taken out of public view, and punish the government just the same as if direct evidence had been exposed. To overcome this problem, the government employs a *cover story*—taking a public action which it privately believes to be ineffective, alongside a more effective but more controversial covert action, in the hopes that the audience will attribute any observed policy success to the observed public action. The cover story succeeds if it mitigates suspicion enough for the

government to avoid backlash from the audience.

In the second part of our analysis, we trace the cover story mechanism 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, in contrast, highlights a series of highly publicized actions by the U.S.—including shipping embargoes, and sanctions and protests registered through the Organization of American States (OAS)—which we argue cannot be fully explained by the administration's desire to use all available means to advance their objective. Rather, we propose that these actions are best understood as part of a cover story strategy. We show that the administration and CIA planners expressed concerns that audiences in Latin America would blame the US for Arbenz's removal even in the absence of direct evidence. We further show that after the mission was complete, the US government drew attention to their public actions in order to disclaim responsibility for covert action—and that observers at the time found the cover story to be convincing.

Third, we highlight surprising implications of our theory for research into the efficacy of a variety of overt instruments of foreign policy, such as sanctions (Marinov, 2005; Davis, 2023), targeted strikes (Kreps and Fuhrmann, 2011; Dell and Querubin, 2018; Allen and Martinez Machain, 2018), arms control agreements (Fuhrmann and Lupu, 2016; Coe and Vaynman, 2020), and public diplomatic statements (McManus, 2017, 2018). By highlighting the strategic interdependence between the use of covert and overt action, our analysis suggests that empirical research in this area faces inferential challenges not previously appreciated: these studies may substantially underestimate *or* overestimate the effectiveness of overt action, depending on the underlying level of transparency and thus the direction of confounding due to unobserved covert action. We derive an observable implication of our model that is integral to this underlying concern: covert and overt action are negatively correlated when transparency is high, but positively correlated when transparency is low. A descriptive quantitative analysis of U.S. foreign interventions during the Cold War period provides evidence consistent with this prediction.

Finally, we argue and provide evidence for the generality of our novel cover story mechanism across diverse political contexts. We present two shorter empirical vignettes that illustrate the theory's applicability to a case of economic negotiations between Australia and East Timor over oil concessions in the Timor Gap, and to the resolution of the Cuban Missile Crisis. This extends the theory's domain to several different policy objectives (regime change, commercial agreements, and nuclear weapons postures), several kinds of secretive policy actions (covert operations, espionage, and private diplomacy), and across multiple state actors.

Overall, this study enriches our understanding of many coercive practices with ambiguous attribution (Baliga, Bueno de Mesquita and Wolitzky, 2020)—including secret proliferation (Debs and Monteiro, 2014), rogue state management (Coe, 2018), cyber conflict (Axelrod and Iliev, 2014), and election meddling (Levin, 2021)—by demonstrating how the (potentially concurrent) use of overt action can complicate attribution in ways not previously considered. Our analysis also contributes to the broader 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; Rauchhaus, 2009; Biddle, Macdonald and Baker, 2018). We similarly demonstrate the value of this framework in explaining how leaders can both be disciplined by the threat of punishments imposed by foreign 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, in isolation or in combination—which likewise yields novel insights into counterintuitive governing behavior: explaining why leaders implement, and broadly publicize, ineffective and costly policies.

1 Secrecy and Plausible Deniability

We examine a setting where political decision-makers desire 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/economic pressure vs. paramilitary operations or assassinations). If the government finds it infeasible to achieve the objective via the less controversial means, it may secretly pursue the more controversial means, and attempt to conceal its actions from the audience.

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 immigration at the southern border, but they do not want the government to achieve this goal by locking children in cages. The first Trump administration initially sought to conceal its policy of family separation from the U.S. and foreign publics, and succeeded in keeping the policy secret for several months (Horowitz, 2021); when revealed, it invited widespread 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). The administration thus 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. 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 the controversial research practices that contributed to breakthroughs (ACHRE, 1996).

While the policy domain is potentially broad, existing research on secret government policy is primarily advanced by scholars of security and conflict studies, with a particular focus on covert intervention (Spaniel and Poznansky, 2018; Poznansky, 2020; Carnegie, 2021). Thus for concreteness, we characterize the government in our theory as an "Intervener" that seeks to influence political

developments in a foreign country, and the secret policies they pursue as a controversial covert operation. The Intervener worries that an exposed covert action will tarnish her reputation (Joseph and Poznansky, 2018; Myrick, 2020; Bloch and McManus, 2024), causing an audience to impose some form of punishment.

The theoretical model is largely agnostic as to who the audience is and why they find a policy to be objectionable. Previous studies of covert action have considered an international community who cares if the Intervener violates international laws and norms (Poznansky, 2025; Bull, 2002); Congress or other political elites who care that a president does not exploit covert action to violate US laws or avoid institutional checks (Smith, 2019; Colaresi, 2012); or the press and the broader public, who care about effective leadership or principles of transparency and integrity in government (Spaniel and Poznansky, 2018).[2] We develop an abstract theoretical model that can incorporate this variety of substantive considerations. We assume that the audience generally deems the covert action—and the Intervener who takes it—to be "unscrupulous", and in conflict with principles that the audience values.

Consistent with existing research, we begin from the assumption that a government pursuing a covert intervention is concerned with the dual objectives of achieving a successful policy outcome, and maintaining *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. Colaresi (2012) studies retrospective Congressional investigations, with a focus on the direct evidence that they uncover. Smith (2019) notes that "without evidence of particular operations, reporters are reluctant to cover news stories." 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

[2] A study by Myrick (2020) finds evidence in support of the claim that the U.S. public is generally opposed to their own leader's covert actions, all else equal. Notably, the author uncovers a secrecy penalty while holding the actions themselves fixed. We consider actions pursued covertly which are inherently more objectionable than actions pursued openly.

costs arising from inference or speculation on the part of the audience.[3]

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, the likelihood that the outcome 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 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.

[3]We use "circumstantial evidence" differently from Canfil (2022). He refers 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.

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

Figure 1: Game Sequence

2 A Model of Covert Action and Cover Stories

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 advisers. The audience can represent any actor who seeks to minimize the Intervener's use of unscrupulous covert actions. As discussed, this might include the Intervener's own electorate, Congress, or mass publics or political elites across different foreign countries.

The game sequence is presented in Figure 1. We discuss each step in turn.

Leader types. To incorporate the reputational considerations discussed in the previous section, we assume the leader has a privately known type, $\theta \in \{0, 1\}$, which determines how intrinsically costly the leader finds covert action to be: $\theta = 1$ denotes a "scrupulous" type, and $\theta = 0$ denotes an "unscrupulous" type.[4] In the first step of the game, this type is drawn by nature and observed privately by the leader; the audience holds a prior belief that $Pr(\theta = 1) = \pi \in (\frac{1}{2}, 1)$.

[4]Our core results can be obtained from a more abstract alternative model without heterogeneous leader types (that is, a model of "pure moral hazard"). We discuss further in Section 2.4.

Policy options. The leader has two distinct policy levers available: a public (or "overt") action $a_p \in \{0, 1\}$, which is observed directly by the audience; and a covert (or "secret") action $a_c \in \{0, 1\}$, which is only observed by the audience with some probability, as we discuss below. The leader can enact either one, both, or neither of these policy levers. Referring back to our opening anecdote for concreteness, a_p can represent the Eisenhower administration's imposition of economic pressure on Cuba through 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.

Before L decides which policies to authorize, in step 2, L receives private information about each of the policy options. First, L learns about the effectiveness of public action, which we denote by the state variable $\omega \in \{0, 1\}$: the leader observes ω privately, and the audience holds a prior belief that $Pr(\omega = 1) = \tau \in (0, 1)$. Second, the leader learns the cost k_c of covert action,[5] which is drawn from a type-specific distribution $F^\theta(x) = Pr(k_c \leq x; \theta)$. We assume that F^0 is continuously differentiable with support $[\underline{k_c}, \overline{k_c}]$, and that F^1 has a lower bound of $(1 + \beta)$; this restriction on F^1 means that covert action is always prohibitively costly for the scrupulous leader, as we discuss below. The leader observes k_c directly, while the audience only knows its distribution.

Substantively, we can interpret step 2 as the leader's advisers presenting her with a private briefing about the policy options available to her, and their best assessments of the relative benefits and drawbacks of each. Returning to the context of Eisenhower's Castro policy, now-declassified documents indicate that the CIA privately briefed White House officials on the efficacy of the sugar quota and oil embargoes, estimating that the Soviets would mitigate the impact of both.[6] 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 model setup 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.[7] Indeed, one Cuban exile leader—while dismissing the possibility of any

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

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

[7] It is intuitive to assume that k_c is private information, as the covert action itself is taken in secret. In our analysis, if the covert action is exposed, it makes no difference whether or not the audience also observes k_c.

sort of armed invasion—stated publicly in early January 1961 that Castro was likely to fall within three months "in view of the economic paralysis and growing discontent" among the Cuban people (Brewer, 1961*a*).

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

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

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

where α_0 denotes the probability of policy success due to exogenous factors, or random luck. We assume $\alpha_0 \leq \alpha_p^0 < \alpha_p^1 < 1$: public action (weakly) increases the probability of success in either state ($\alpha_p^{\omega} \geq \alpha_0$); it is more effective when $\omega = 1$ than when $\omega = 0$ ($\alpha_p^0 < \alpha_p^1$); but it never guarantees success ($\alpha_p^{\omega} < 1$). Likewise, we assume $\alpha_0 < \alpha_c < 1$ (but impose no restriction on the relative efficacy of covert vs. public action). When covert and public action are taken simultaneously, the probability of success is[8]

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

Covert revelation. When the leader takes covert action, she initially does so in secret, but direct evidence of the covert action may later be inadvertently revealed (step 5 of the game sequence). Let $z \in \{0,1\}$ denote whether covert action is exposed, with

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

[8]For intuition, we can think of this as the complement of the probability of failure, modeled as the joint probability of both covert and public action failing independently: $Pr(y = 0|a_p = a_c = 1, \omega) = (1 - \alpha_p^{\omega})(1 - \alpha_c)$. See Spaniel and Poznansky (2018) for a similar modeling approach.

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 $\lambda \in (0, 1)$. We refer to λ as the level of *transparency* in the policymaking environment.

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

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

The substantive interpretation of the audience's action can vary depending on the context. For a domestic voter, r can represent the choice of whether to support the incumbent leader against her electoral challenger. For a domestic legislature, it can represent the choice to to impose some form of punishment on the executive, for instance in the form of withholding funding of its policy priorities, or investigating or legislatively curtailing its authority (Colaresi, 2012; Spaniel and Poznansky, 2018). For foreign audiences, r can represent the choice over whether to cooperate with L on future foreign policy initiatives, or to withdraw from L's bloc or alliance system more broadly (Poznansky, 2025). In each case, the audience prefers to "reward" the leader if and only if she is scrupulous, but faces uncertainty as to her true type. We assume $\pi > \frac{1}{2}$, meaning that leader enjoys a "presumption of innocence"; the audience is disinclined to punish the leader based on their prior beliefs of her type, but may be swayed toward punishment on the basis of direct or circumstantial evidence.

Leader's payoff. The leader's payoff is

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

Both leader types enjoy a benefit normalized to one for a successful policy outcome (and zero for failure); they receive a political or reputational benefit of β when rewarded by the audience (with the "penalty" of punishment normalized to zero); and they pay direct costs k_p and k_c for taking public action and covert action, respectively. As stated above, the factor that distinguishes the two

types of leader is the distribution from which their covert action cost k_c is drawn: for scrupulous leaders, we assume $k_c > 1 + \beta$, which means that they always find covert action to be prohibitively costly. Unscrupulous leaders are less intrinsically opposed to taking covert action; whether or not they do so depends on the realization of k_c, and the incentive scheme created by the audience's endogenous punishment/reward strategy.

Comment on endogenous plausible deniability. Past models of plausible deniability and covert action assume that the leader is automatically punished—implicitly, by some non-strategic audience—whenever direct evidence of covert action is revealed. In contrast, we assume that a strategic audience has incentives to punish the kinds of leaders whom they believe are willing to pursue unscrupulous covert action, and to reward the kinds of leaders who are not. To preview what will come, the core question guiding the audience's punishment strategy is the extent to which they believe the leader to be unscrupulous. This belief depends on three sets of factors: first, A's prior beliefs about the pieces of information which the leader observes privately (the leader's type, and the costs and benefits of the available policy levers); second, the additional information the audience observes over the course of the game (the leader's choice of overt action a_p, the policy outcome y, and the revelation (or not) of covert action z); and finally, the audience's conjecture of the leader's strategy (which is correct in equilibrium). By incorporating these considerations into the audience's strategy, we demonstrate how the leader's challenge of plausible deniability becomes much more complex than a simple concern over operational security and minimizing the risk of direct exposure.[9]

2.1 Technical Preliminaries

We introduce two substantively motivated assumptions. First, we impose the following parameter restrictions:

Assumption 1 (Parameter restrictions) *Throughout the analysis, assume the following:*

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

(ii) $\alpha_0 < \min\left\{\alpha_c(1-\lambda), \alpha_p^1\alpha_c\right\}$

[9] A related advantage is that we do not mechanically assume that overt action arouses suspicion. Rather, we structure the model to demonstrate how audience suspicion arises endogenously from the observation of overt action.

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

(iv) π is in an intermediate range (with the bounds defined in the appendix)

Collectively, these assumptions ensure that the leader faces the strategic problem that motivates us: maintaining plausible deniability for objectionable covert action in the face of strategic inferences. We explain each point in greater depth in Appendix 7.1.

Second, we focus attention on a set of substantively appealing equilibria. We define the equilibria of interest as follows:

Definition 1 *A **responsive equilibrium (RE)** is an equilibrium in which the scrupulous leader takes public action if and only if the direct policy benefits outweigh the direct costs: that is, she plays $a_p = \omega$.*

Any behavior by the scrupulous leader which does not satisfy this condition would be a form of "pandering", phenomenon which has been studied thoroughly in previous work (Canes-Wrone, Herron and Shotts, 2001; Maskin and Tirole, 2004) but which is not the substantive focus of our analysis.[10] Later in the analysis (following the presentation of Corollary 3), we explain the distinction between pandering and the novel mechanism that we develop.

We can first establish the general existence of these non-pandering equilibria:

Proposition 1 (RE Existence) *A responsive equilibrium always exists.*

While the RE we characterize below are not necessarily unique, the following result provides a justification for focusing our analysis on the RE even when other equilibria can be supported.[11]

Corollary 1 (RE Optimality) *Among all equilibria, the RE yields the best policy payoff for the scrupulous leader. If α_0 is low, the RE yields the best overall payoff for the scrupulous leader.*

Throughout the main text, we will impose the following restriction:

Assumption 2 *Restrict attention to responsive equilibria.*

This restriction is made simply for ease of exposition; in the appendix, we present an analogue of our main result (specifically, Proposition 3) which does not invoke this equilibrium selection rule.

[10] A similar restriction could be accomplished more simply by treating the scrupulous leader as a "behavioral type" who always governs in line with the audience's interests; see Ashworth and Bueno de Mesquita (2014)

[11] A separate justification for focusing on this equilibrium follows from Proposition 4 in the appendix, under the pure moral hazard setting.

2.2 Analysis

When a leader's most effective policy option is a controversial covert action, how do they navigate the dual objectives of achieving a successful policy outcome while maintaining plausible deniability for the actions they took to achieve it? Our analysis reveals that the leader's optimal strategy depends on transparency, with novel behavior emerging when transparency is low.

We begin by defining our novel cover story mechanism:

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

Our model is structured so that the leader has two possible incentives for taking public action: increasing the probability of achieving a successful policy outcome, and maintaining a favorable reputation with the audience. Our definition of a cover story applies to situations in which public action cannot be justified by the first incentive alone; formally, this is the condition that $\omega = 0$, in which case the direct costs of public action outweigh the direct benefits,

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

as per Assumption 1 (iii).

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

Proposition 2 *There exist thresholds λ^* and λ^{**} such that:*

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

- *If transparency is low ($\lambda < \lambda^*$), the leader uses a cover story with positive probability; that is, all equilibria are CSE.*

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

2.3 The disciplining effect of high transparency

In general, from A's utility function (3), we can see that the audience's equilibrium strategy will depend on their posterior belief of the leader's quality. Given the observed history of the game $h = (a_p, y, z) \in \{0,1\}^3$, the audience forms a belief $\mu^h = Pr(\theta = 1|h)$. They will prefer to reward the leader ($r = 1$) if $\mu^h > \frac{1}{2}$, and punish ($r = 0$) if $\mu^h < \frac{1}{2}$.[12]

Recall that the scrupulous leader never takes covert action. It follows that upon observing direct evidence of covert action ($z = 1$), the audience will draw the most negative possible inference of the leader's scrupulousness ($\mu^{a_p,y,z=1} = 0$), and will punish the leader accordingly. Thus when transparency is high ($\lambda \geq \lambda^{**}$), the risk of direct exposure and ensuing punishment can effectively discipline the unscrupulous leader into never taking covert action.[13] Under this condition, both leader types exhibit the same behavior, taking only public action whenever the state is favorable ($a = (a_p = 1, a_c = 0)$ when $\omega = 1$), and otherwise doing nothing ($a = (0,0)$ when $\omega = 0$).

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

[12] It is without loss of generality to fix the "punishment threshold" at $\frac{1}{2}$; all that matters is the difference between this threshold and the prior belief π.

[13] Note that for some parameter values, this condition of high transparency may not exist. For instance, if covert action is very effective ($\alpha_c - \alpha_0$ is large) and the lower bound on its direct cost ($\underline{k_c}$) is very low, then there is no $\lambda \in [0,1]$ for which the unscrupulous leader can be completely deterred from ever taking covert action.

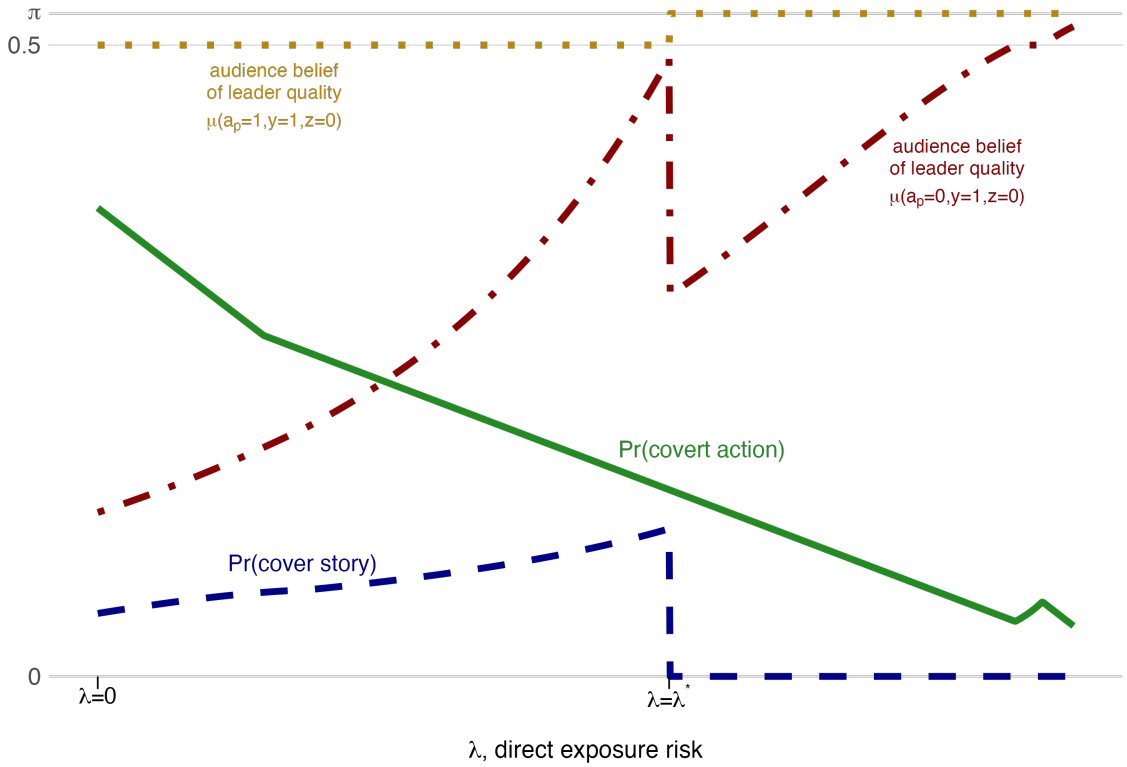
2.3.1 The problem of unexplained success

When transparency falls into the intermediate range (λ falls below λ^{**}), the leader's incentives begin to change, and covert action (sometimes) becomes a worthwhile gamble: if the direct costs (k_c) are low, the leader will accept the risk of exposure (and accompanying reputational harm) in exchange for increasing her chances of a successful policy outcome. This, in turn, complicates the audience's inference. Upon observing an unexplained success, the audience infers that one of two things must have happened: either the leader took no action, and the success arose due to random luck; or the leader took covert action, with no direct evidence of her action coming to light. The relative weight that the audience assigns to each possibility depends on the level of transparency. As transparency decreases, the leader becomes more likely to take covert action, and the audience becomes less likely to observe covert action if it is taken—both of which contribute to a less favorable inference following unexplained success.

This shift in leader strategy and audience beliefs is visualized in Figure 2. On the far righthand side of the figure, at the highest level of transparency, the leader uses covert action only infrequently; after observing an unexplained success, the audience is willing to believe that the outcome is attributable to random luck, and they refrain from punishing the leader as a result. This favorable inference by the audience can only be sustained up to a point, however. As transparency decreases, so too does the audience's belief of the leader's scrupulousness: 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 an unexplained success. Eventually, the audience's posterior belief $\mu^{a_p=0,y=1,z=0}$ (the red dot-dashed line in the figure) falls below $\frac{1}{2}$, and they fully punish the leader purely on the basis of circumstantial evidence.

At the same time, in the intermediate range of transparency ($\lambda^* < \lambda < \lambda^{**}$), the audience's belief of the leader's scrupulousness is more favorable after seeing her take public action (and, of course, when no direct evidence of covert action is exposed). This follows simply from their (correct) conjecture of the leader's equilibrium strategy: the leader only takes public action when it is effective; and the use of effective public action makes covert action largely redundant, and

Figure 2: 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. Note that under the paramater values used to construct this figure, $\lambda^{**} > 1$, so the figure only depicts equilibrium behavior within the "low" and "intermediate" ranges of transparency.

unnecessarily risky.[14] The fact that the leader took public action thus serves as an informative signal that the leader is not likely to have taken covert action, giving the audience little reason to believe that L is unscrupulous. This is visualized in the gold dotted line in the figure remaining flat at π over the range of $\lambda > \lambda^*$. As a result, the audience rewards the leader when they observe public action under intermediate (or high) transparency.

2.3.2 The value of the cover story

The two aforementioned features of the audience's strategy—punishing unexplained success, but rewarding public action—provide the rationale for the leader's use of a cover story. Consider the

[14] For $\lambda \in (\lambda^*, \lambda^{**})$, the unscrupulous leader may sometimes take covert action and public action simultaneously when $\omega = 1$ (that is, when k_c is especially low); the lower bound on π, given in Assumption 1 (iv), implies that this probability is low enough that the audience optimally refrains from punishment after observing public action (i.e. that $\mu^{1y0} > \frac{1}{2}$).

unscrupulous leader's evaluation of her policy options when she (privately) expects public action to be ineffective (i.e. when $\omega = 0$), but a relatively low-cost covert action is available (k_c low). She could pursue the covert action alone, accepting that if it succeeds, she will face severe backlash from the audience whether or not direct evidence comes to light. Alternatively, in addition to pursuing the covert action, she could also take the ineffective public action, claiming that she actually expects it be effective. If the successful outcome is achieved, and direct evidence of covert action remains unexposed, she can point to the public action as the cause of the policy successful and hope that her audience is willing to accept that attribution.

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

Corollary 2 (CSE comparative statics) *The leader uses a cover story when transparency is low, $\lambda \leq \lambda^*$. The threshold λ^* is:*

- *increasing in the effectiveness of covert action, α_c;*
- *decreasing in the direct cost of public action, k_p;*
- *and, if α_0 is low, λ^* is increasing in the leader's value for audience approval, β.*

The leader's use of a cover story entails a tradeoff between the direct cost of public action (k_p), and the reputational benefits bestowed by the audience (β).[15] The cover story only serves to *improve* the leader's reputational payoff (relative to taking covert action on its own) if (i) covert action is not directly exposed (which occurs with probability $1 - \lambda$), and (ii) the policy succeeds (probability α_c), since the audience has little reason to suspect covert action given an unsuccessful outcome. Thus the net benefit of a cover story is increasing in α_c and in β, and decreasing in k_p and in λ.

It is worth pausing to clarify exactly how the leader benefits from using a cover story. Technically, our model setup assumes that the audience observes a_p, y, and z simultaneously, and chooses whether to punish or reward the leader given all three pieces of information. As a practical matter,

[15] There is also the slight benefit of an increase in the probability of policy success, by $\alpha_p^0(1-\alpha_c)$. The lower bound on k_p given in Assumption 1 (iii) implies that this policy benefit alone would not justify using ineffective public action, without an additional reputational benefit.

however, there is a plausible sequencing of information, wherein the audience first observes the leader's action, and only after some time observes the policy outcome and the revelation of covert action (or lack thereof). Considering the audience's interim beliefs—after observing the leader's action but before observing the outcome—can provide some substantive insight on what the cover story does and does not accomplish for the leader.

Corollary 3 (Cover Stories and Scrutiny) *In any CSE, the audience's interim beliefs of the leader's scrupulousness after observing public action (but before observing the outcome, or any revelation of covert action) are strictly less favorable than their interim beliefs after observing no public action.*

As suggested in the anecdote of Eisenhower's Castro policy, 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 public view. Indeed, Eisenhower's announcement of severing diplomatic relations led to immediate speculation—both by concerned pro-Castro groups, and by enthusiastic Cuban exiles—that the United States was planning a covert intervention to overthrow the Castro regime (Times, 1961; Brewer, 1961*b*). Our analysis demonstrates that, despite raising suspicion in the short term, cover stories ultimately help leaders provide a long-term answer to the question of how events turned in their favor without their having resorted to unscrupulous covert action. 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.

This finding also helps to clarify the distinction between cover stories and "pandering" (Canes-Wrone, Herron and Shotts, 2001; Maskin and Tirole, 2004). In a pandering equilibrium, the leader would make a policy choice that she believes is in neither her nor the audience's policy interests (playing $a_p \neq \omega$, in our setting) because doing so is "popular"—that is, because the audience expects the scrupulous leader to behave that way, and thus rewards that behavior and punishes any other behavior. In our cover story equilibrium, in contrast, Corollary 3 says that taking the public action (as part of the cover story) is actually *unpopular*, relative to doing nothing at all. The leader relies on a cover story because it is a less-bad option than achieving her desired policy

outcome through objectionable means with no alternative explanation.

2.4 Generality of the mechanism

Here we briefly highlight three points pertaining to the generality of the model's key insights.

Equilibrium selection. We focused on Responsive Equilibria (RE) for clarity of exposition. However, as we show in the appendix, the relationship between transparency and cover stories does not depend on this restriction. In general, we show that Cover Story Equilibria exist if and only if transparency is below a threshold λ^* (Proposition 3). We further note that when $\lambda < \lambda^*$, the only non-CSE equilibria that exist involve the audience fully punishing the leader whenever they observe public action (even in the absence of any direct evidence of covert action). We find this pattern of behavior to be substantively implausible, as it collapses any distinction between the reputational implications of, e.g., authorizing economic sanctions vs. assassination attempts against foreign leaders.

Homogeneous leader types. We focused on the distinction between scrupulous vs. unscrupulous leaders, and the audience's challenge of distinguishing between them. This follows a long tradition in the political agency literature, which holds that prospective concerns over screening agents by quality will generally dominate any incentive to impose punishments on a purely retrospective basis (Fearon, 1999; Ashworth, 2012). However, in Appendix 7.4 (Proposition 4), we show that our novel cover story mechanism does not require heterogeneous leader types. Rather, it can still emerge in a model of "pure moral hazard", in which all leaders are commonly known to be unscrupulous, and the audience is simply trying to design an incentive scheme that minimizes the leader's use of covert action.

Substantive policy domain. For concreteness, our presentation of the formal model focused on a setting of foreign intervention involving overt and covert policy options. This fits the in-depth case study we presented in Section 3. However, the core logic of a cover story depends on only a few key representational features: a policymaker can pursue a policy objective through actions that are either public and acceptable, or secret and objectionable; and she can be held accountable by an audience who wants to prevent her from using the objectionable means. As discussed above, we

believe these assumptions can characterize a much broader set of substantive policy contexts, such as immigration enforcement or scientific advancement. In Section 5, we further support this claim with two shorter case studies focused on issues of commercial disputes over natural resources, and diplomatic negotiations over nuclear weapons postures.

2.5 Empirical Implications for Covert Action Research

One implication of our theory, relevant to all historical research into covert action, highlights strategic complications that mission planners must navigate to achieve plausible deniability:

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

This implication highlights the general inevitability of concerns over plausible deniability in the conduct of covert operations. The particular nature of the concern, however, will differ across contexts. When transparency is higher, mission planners will primarily focus on maintaining tight operational controls that minimize the risk of direct exposure. When transparency is lower, they will be more concerned with suffering reputational harm on the basis of circumstantial evidence alone.

In light of these concerns, when Interveners use covert action under conditions of low transparency (and the other conditions outlined in Corollary 2), we further expect to observe the following:

Implication 2. 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, than the secret policies that the leader hopes to cover up. It should

not be too intrinsically costly (k_p low), and it should be plausible, from the audience's perspective, that the action could significantly contribute to the likelihood of policy success ($\alpha_p^1 > \alpha_0$).

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

Implication 4. The audience should be convinced by the cover story, and willing to attribute the observed outcome to the observed public actions—at least to a sufficient degree that they are willing to refrain from punishing the Intervener.

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

[16] Quoted in Schmitz (1999, 179)

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, 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 3.3 details the overt actions the administration took, and considers how well existing accounts can explain these actions. Section 3.4 demonstrates that the cover story mechanism can explain these overt actions, highlighting the Eisenhower adminsitration's concerns over strategic inferences, as well as audience receptions of the cover story. 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 20th 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". 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 concern 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,

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

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 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]

The necessary conditions for the cover story equilibrium to hold are that the risk of direct exposure (λ) is low; covert action is relatively effective (high α_c); the leader's reputational concern (β) is relatively high; and the direct cost of public action (κ_p) is moderately low (but not zero). At this time, covert regime change operations were in their infancy, but the recent success of the same playbook to oust Mossadeq in Iran undetected gave the Administration confidence that both the

[18] As another potential case, many features of Eisenhower's reasoning to oust Castro fit our cover story equilibrium. But this 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.

proposed plan was their best chance of success (relatively high α_c), and that they could avoid direct evidence via tight operational controls (low λ) (Cullather, 2006, 7). 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 β. 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 (κ_p moderately low).

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 seek to divert public attention away from Guatemala as CIA officers were in the field, in order to minimize the risk of direct exposure. This is not what we observe. 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 up their overt policies. On May 15, a freighter carrying weapons that Arbenz had purchased from Czechoslovakia landed in Guatemala (Immerman (1982, 155); Schlesinger and Kinzer (1982, 147)). Arbenz had hoped to keep the shipment 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 Guatemalan government. This was at Eisenhower's direction, who instructed his diplomats that "By every proper and effective means we should demonstrate to the courageous elements within Guatemala who are trying to purge their government of its communist elements that they have the sympathy and support of. . . the 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 U.S. concerns over 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 direct 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 whether Eisenhower's desire to deploy all effective policy instruments can fully account for the overt action that we observe. If so, 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*g*). 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*e*). He went on to warn: "There is not the slightest doubt that if the operation is carried through many Latin Americans will see in it the hand of the 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*d*). 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 needed 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 over U.S. involvement, Wisner argued that "it might be a good idea to cry wolf several times before D-Day" (CIA, 1954*f*). 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*c*).

3.4.2 Employing the cover story

Our third empirical prediction is that administration officials should seek to connect the observed outcomes to the public actions, so as to disclaim any covert regime change operations. In this case, we find evidence of some elites privy to the operation explicitly articulating a cover-story motivation in 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.*[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.

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 involvement. An NSC report, later released to the press, argues that the US contributed to pressuring Arbenz to resign through several overt actions. The report states that "The Organization of American States was used as a means of achieving our objectives in the case of communist intervention in Guatemala"; following the discovery of the arms shipment, "the

[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*a*); emphasis added.

United States initiated consultations with all Latin American Governments, except Guatemala." When the Arbenz Government appealed to the OAS to accuse its neighbors of aggression, the U.S. "took the position that the Organization of American States was ready, willing and competent to respond to the appeal"; but before an investigation could proceed, "the new government of that country indicated that the controversy requiring the investigation had ceased to exist" (CIA, 1955). By highlighting U.S. diplomatic actions taken out in the open, the report implies that other, unscrupulous means of achieving the same objectives were not being pursued.

3.4.3 Public reception of the cover story

Finally, while we expect that observers will naturally be suspicious of U.S. covert activity, their suspicions will be offset in part by the lack of direct evidence, and in part by their belief that 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 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)—a U.S.-based historian of Latin America with no governmental affiliation—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 finds:

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

He separately considers whether 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 concludes: "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.[22] 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 have been successful, again speaking to the perceived efficacy of U.S. overt policies (p. 804–805). 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.

In addition to Taylor's in-depth account, similar interpretations of the events are offered by a broader range of journalists. For example, Harsch (1954) offers a more positive view on potential U.S. intervention in Guatemala: "If there were no native revolutionary movement to encourage and support, then some other ... remedy would have to be found." But he believes that covert action ultimately proved unnecessary because, "Fortunately, there was a bona fide native movement;

[22] 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).

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. intervention in Latin America. For example, Reston (1954) openly speculates that the CIA was involved in Guatemala, but stops short of asserting that they played a directly role in the coup—merely noting instead that the CIA was integral in uncovering the weapons cache, and exploiting that episode to foster anti-Arbenz sentiment.

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 criticized for overt actions that were viewed as intrusive meddling against a democratically 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, we theorize that cover stories do not conclusively convince the target audience that covert action was not taken. Rather, they offset suspicion only enough for the Intervener to avoid backlash. Indeed, some observers speculated about 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. But the suspicion was not enough to cause the reputational harm that had worried Eisenhower at the outset, and which we discuss at the beginning of the case. Indeed, the lack of international domestic backlash is part of the reason that historians consider PBSUCCESS a successful case of covert regime change (Immerman, 1982; Schmitz, 1999).

In sum, the evidence shows that before PBSUCCESS 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

We now address two potential concerns that might be raised by researchers who study PBSUCCESS from a broader range of historical perspectives.

First, we recognize that the cover story was just one mechanism by which mission planners sought to manage strategic inferences. 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.[23] The CIA also armed the coup-plotters with weapons it purchased from the Dominican Republic to implicate them (CIA, 1954*b*).

We do not claim that avoiding strategic inferences is uni-causal. These alternative methods are consistent with our overall theory and speak to the importance of studying the tensions between avoiding strategic inferences and direct evidence more broadly. 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."[24]

Second, one might wonder if Eisenhower engaged others at the OAS to offset backlash in the event that PBSUCCESS was exposed. 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

[23]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 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."

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

or make inflammatory statements against Arbenz outside of the OAS meetings.

4 Implications for Empirical Studies of Overt Action

Our analysis thus far has primarily focused on leaders' decisions over when to use controversial covert actions, and how they can evade accountability for doing so. Beyond speaking to this motivating question, the theory we develop also carries important implications for research aimed at evaluating the efficacy of a wide range of overt policy instruments—including sanctions (Marinov, 2005; Davis, 2023), targeted strikes (Kreps and Fuhrmann, 2011; Dell and Querubin, 2018; Allen and Martinez Machain, 2018), public diplomatic statements (McManus, 2017, 2018), and arms control agreements (Fuhrmann and Lupu, 2016; Coe and Vaynman, 2020). In this section, we demonstrate how the strategic interdependence between covert and overt action can pose major inferential challenges for empirical studies of overt action. Without fully accounting for the occurrence of unobserved covert action, empirical analyses can substantially underestimate *or* overestimate the efficacy of overt action, depending on the underlying level of transparency in the policymaking environment.

From the theoretical model presented in Section 2, we can derive the following result:

Corollary 4 (Correlation between public and covert action)

- *If transparency is sufficiently high, covert action and public action are negatively correlated.*
- *If transparency is sufficiently low, covert action and public action may be positively correlated.*

The first claim reflects a standard view in the literature: covert action is more likely to be taken in situations where overt action is infeasible or undesirable, and vice-versa (Downes and Lilley, 2010; McManus and Yarhi-Milo, 2017; Joseph and Poznansky, 2018; Smith, 2019). In our model, it follows simply from the technological assumption that covert and public action are substitutes in the policy production function; that is, because we assume $E[y|a_p = a_c = 1, \omega] = \alpha^{\omega}_{pc} = \alpha^{\omega}_{p} + (1 - \alpha^{\omega}_{p})\alpha_c$, it follows that

$$\begin{pmatrix} E[y|a_p = 1, a_c = 1, \omega] \\ -E[y|a_p = 0, a_c = 1, \omega] \end{pmatrix} < \begin{pmatrix} E[y|a_p = 1, a_c = 0, \omega] \\ -E[y|a_p = 0, a_c = 0, \omega] \end{pmatrix}$$

for $\omega = 0, 1$. This expression says that public action has a greater marginal effect on the outcome when covert action is not taken vs. when covert action is taken; rearranging, we obtain an analogous claim for the marginal effect of covert action.

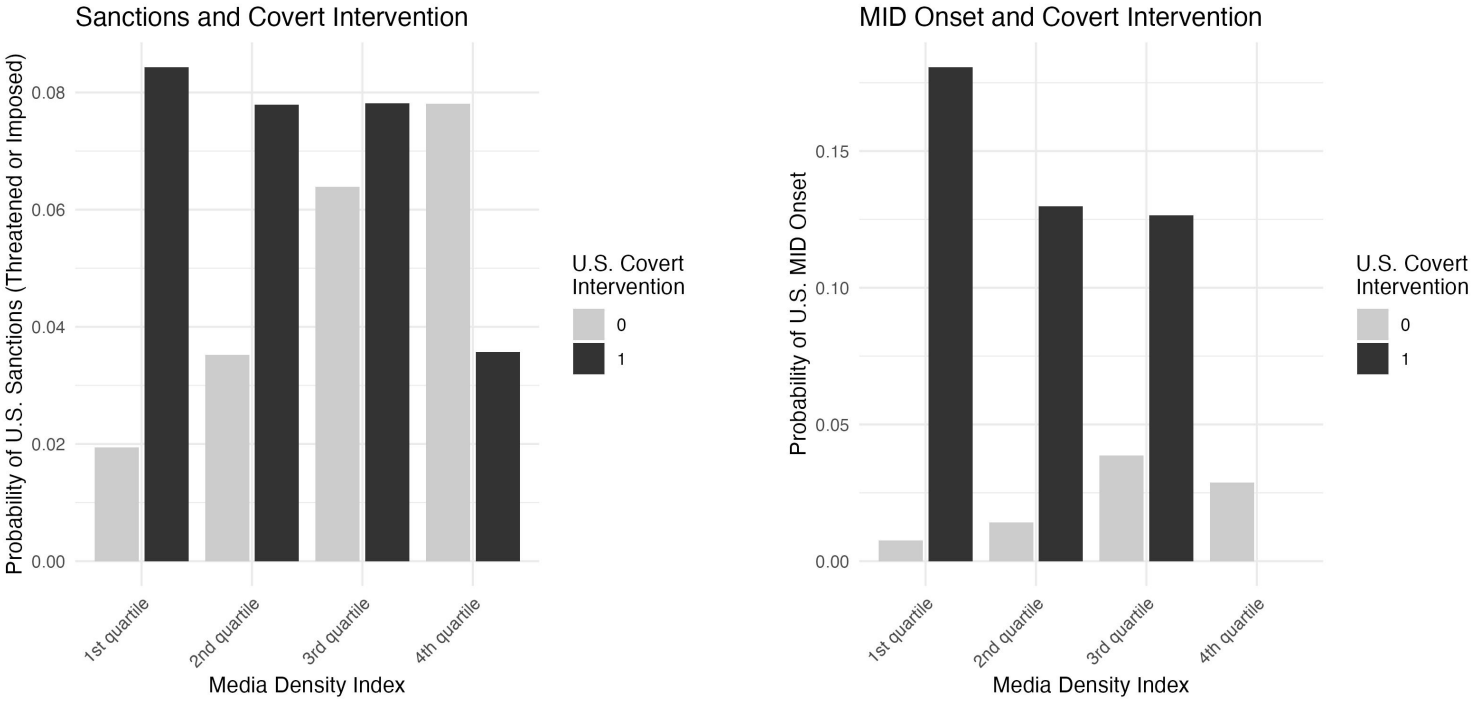
If the audience were not capable of making strategic inferences (or if the leader was not concerned with audience approval), this substitutability assumption would imply a negative correlation between the use of covert and public action across the entire parameter space. By allowing for a richer strategic interaction between the leader and audience, however, we uncover a different relationship under conditions of low transparency. The leader's use of ineffective overt action as a cover story for effective covert action can result in a positive correlation between the two forms of action.

To be clear, our theory does not definitively predict that covert and public action *must be* positively correlated under low transparency—only that they *can be* (depending on other parameter values, and depending on which of possibly multiple equilibria are selected). Yet the possibility of a positive correlation, conditional on the underlying level of transparency, sets our model apart from standard arguments in the literature. Evidence of such a relationship in the empirical record would be consistent with our model, but is not predicted by existing theories.

As a first pass at evaluating the conditional relationship between covert and public means of coercive foreign policy, we analyze a cross-national dataset, drawing on measures used in closely related studies to operationalize the theoretical quantities of interest. Specifically, we use the country-year dataset compiled by Smith (2019), which draws on O'Rourke (2018)'s measure of CIA-initiated foreign regime change operations, and Gibler, Miller and Little (2016)'s measure of U.S.-initiated Militarized Interstate Disputes (MIDs), as measures of covert and public actions, respectively. We also incorporate a measure of U.S. threats or impositions of economic sanctions from the TIES dataset (Morgan, Bapat and Kobayashi, 2014) as a separate measure of public action, following previous work examining the impact of sanctions on the survival of targeted leaders (Marinov, 2005; Escribà-Folch and Wright, 2010). Finally, we follow Joseph and Poznansky (2018) in using Warren (2014)'s measure of "media density" to operationalize our concept of transparency, or the risk that evidence of a covert operation is exposed and made public. The sample covers 157 countries from 1947–1989, for a total of 5,080 country-year observations.

We conduct a simple descriptive analysis, separately with each of the two measures of public

Figure 3: Conditional Relationship between Covert and Overt Action

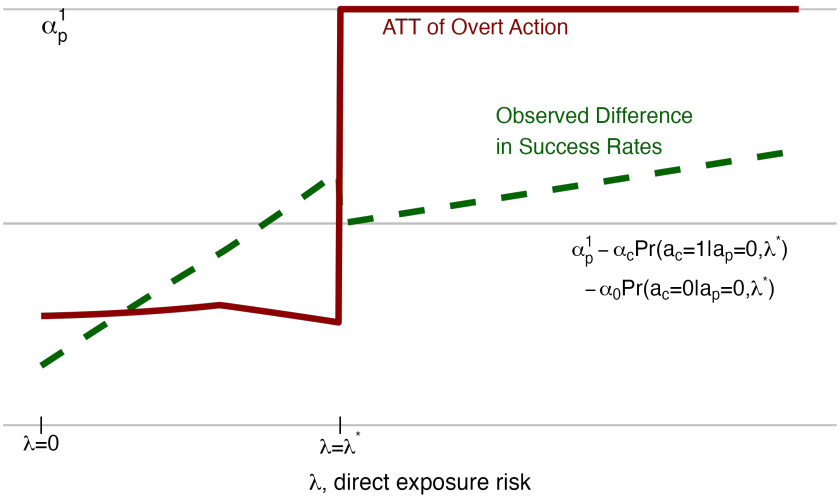


Note: Within each panel, the sample of 5,080 country-year observations (157 countries from 1947–1989) is split into quartiles of Media Density Index. Within each quartile, black bars denote the probability of U.S. overt action (sanctions or MIDs), conditional on there being an ongoing U.S. covert intervention in the country; gray bars denote the probability of U.S. overt action, conditional on no covert intervention.

actions.[25] We divide the sample into quartiles of media density; within each quartile, we report the conditional probabilities of public action being used, conditional on covert action simultaneously being employed or not. The results are reported in Figure 3. First consider the left panel. For the lowest quartile of media density, we find a strong positive correlation between the use of covert interventions, and the threat or imposition of sanctions: sanctions occur in 8% of cases where covert intervention is also being pursued, but in only 2% of cases without covert intervention. At the highest quartile of media density, the relationship is reversed: sanctions are *less* likely to be used when there is a covert intervention occurring (3.5% of cases) than when there is not (8% of cases). A similar but even starker pattern emerges when using MID onsets as the measure of public action in the righthand panel: at the lowest levels of transparency, MIDs are far more likely to occur when covert action is also being pursued than when it is not; but at the highest levels, MIDs

[25]Results are qualitatively similar if we use the onset of "minor" MIDs (which might be better suited to our concept of low-cost, ineffective public actions), rather than all MIDs as we do in the analysis reported in Figure 3.

Figure 4: Treatment effect of overt action vs. observed correlation with policy success



Note: Solid red line denotes the average treatment of the treated (ATT) of overt action on policy success. Dashed green line denotes the observed difference in the probability of success between cases in which overt action is taken vs. cases in which it is not.

and covert action *never* coincide.

In light of these findings, suppose an empirical study seeks to estimate the true effect of some form of overt action. In our model, we can define the Average Treatment Effect among the Treated (ATT) of overt action as follows: among the cases in which overt action is taken, what is the difference between the observed rate of policy success, vs. the counterfactual rate of success if overt action had not been taken. Compare this quantity to the Observed Difference in Means (ODIM) that a naive empirical study might estimate, without accounting for covert action, or for the unobservable state ω: this would be the difference in observed success rates among cases in which overt action is taken, vs. cases in which overt action is not taken.

Remark 1 *The observed difference in mean success rates (ODIM) can be ether positively or negatively biased for the true ATT of overt action, with the magnitude and direction of bias varying with the level of transparency.*

We illustrate this phenomenon in Figure 4. When $\lambda > \lambda^*$, the leader never uses a cover story, so covert and overt action are negatively correlated. The ODIM compares average success rates in cases where overt action is taken (α_p^1), vs. in cases where it is not ($\alpha_c Pr(a_c = 1|a_p = 0) + \alpha_0 Pr(a_c = 0|a_p = 0)$); the latter category includes a substantial portion of cases where unobserved covert action

is being used, which increases the success rate (relative to what it would be if no action of any kind were taken). As a result, the ODIM underestimates the true effect of overt action (where the true ATT is $\alpha_p^1 - \alpha_0$).

When $\lambda < \lambda^*$, and the leader (sometimes) uses a cover story, the bias of the naive estimator becomes more complicated, reflecting a combination of factors. There remains a negative correlation between covert action and *effective* overt action (that is, under the condition that $\omega = 1$). However, we now start to see *ineffective* overt action being used concurrently covert action (when $\omega = 0$). This can lead a researcher to overestimate the average effectiveness of overt action, because the impact of unobserved covert action is being misattributed to the observed overt action.

5 Broadening the Argument

As the final piece of our analysis, we apply our theoretical mechanism to two additional empirical vignettes covering a wider range of substantive policy domains outside of the covert action context. This follows previous work that seeks to demonstrate how a hard-to-observe but general mechanism can provide insight into many substantive contexts (e.g, Coe and Vaynman, 2020). We summarize our cases and their differences in Table 1. While not claiming to provide a comprehensive explanation for either case, we illustrate that our mechanism could plausibly illuminate underappreciated dynamics that arise across diverse policy scenarios of interest to a broader range of scholars.

5.1 The Timor Gap Scandal

When East Timor 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

[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.

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

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)."

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 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. Legal analysts believed that East Timor could accrue substantial concessions if the matter was referred to an international court (King, 2017).

In 2006, the parties signed the Treaty on Certain Maritime Arrangements in the Timor Sea (CMATS). While Australia made some concessions, analysts agree that CMATS substantially favored Australia (Cleary, 2007). It included a 50-50 split on the Sunrise Gap, and a commitment from East Timor that they could not renegotiate for 30 years. 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 Australia 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 policy actions, 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.

5.2 Cuban Missile Crisis

According to public understanding at the time, the Cuban Missile Crisis ended when the USSR withdrew missiles from Cuba in return for a vague commitment from the US not to invade Cuba. It is now well known that the resolution of the crisis is largely attributable to Kennedy's secret committment 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 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 the proximity of US missiles to the Soviet Union justified the Soviet missile deployment to

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

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 (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

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

observers that the Soviets did capitulate for fear of escalation. Indeed, the Administration used their tough public stance to help explain why the Soviets eventually backed down, with Secretary of State Rusk famously stating: "We're eyeball to eyeball, and I think the other fellow just blinked."

6 Discussion

In this paper, we put forward a strategic theory of plausible deniability, focusing on the need for governments to manage reputational concerns arising from circumstantial evidence of their involvement in controversial secret policies. We trace our novel "cover story" mechanism through an in-depth case study of the CIA's Operation PBSUCCESS in Guatemala, and through shorter empirical vignettes of negotiations over oil concessions in the East Timor Gap and the resolution of the Cuban Missile Crisis, demonstrating the breadth of the theory's applicability across substantive domains. With quantitative evidence of U.S. overt and covert interventions throughout the Cold War, we demonstrate how unobserved covert action can pose significant inferential challenges for empirical research evaluating the efficacy of overt instruments of foreign policy.

Our theory holds implications for global security given the re-emergence of great power competition and threats to the Liberal Order. The logic of strategic inferences means that grey zone conflict may play a different role in US relations with Russia and China than existing research would predict. Some existing work emphasizes that states can avoid escalation and retaliation if they can conceal direct evidence of an attack (Carson, 2016; Cormac and Aldrich, 2018; Napier, 2023; Bloch and McManus, 2024). 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 audiences engage in strategic inferences, 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 International 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.

On the other hand, strategic inferences ease concerns that the mere possibility of covert action will degrade international laws and norms (Lake, Martin and Risse, 2021; Farrell and Newman, 2021). There is mounting evidence that violating international laws and norms is costly (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; Poznansky, 2020). Our theory suggests a practical limitation on how frequently states can use covert action to circumvent international responsibilities, especially if they use covert actions repeatedly over time.

Finally, our theory also holds implications for domestic politics given growing mistrust in government. It explains for public accountability activists 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. It also suggests that policymakers can offset widespread conspiratorial beliefs through performative overt policies. Conspiracies often enter public consciousness when there is no plausible explanation for events that the audience knows to be in the government's interest. Cover stories can fill that void.

References

Acharya, Avidit, Matthew Blackwell and Maya Sen. 2016. "Explaining causal findings without bias: Detecting and assessing direct effects." *American Political Science Review* 110(3):512–529.

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

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

Allen, Susan Hannah and Carla Martinez Machain. 2018. "Choosing air strikes." *Journal of Global Security Studies* 3(2):150–162.

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.

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

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

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

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

CIA. 1954*g*. "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.

Coe, Andrew J. and Jane Vaynman. 2020. "Why Arms Control Is So Rare." *American Political Science Review* 114:342–355.

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.

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

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

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

Davis, Jason. 2023. "Targeted Sanctions and Redistribution.". working paper.

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

Dell, Melissa and Pablo Querubin. 2018. "Nation building through foreign intervention: Evidence from discontinuities in military strategies." *The Quarterly Journal of Economics* 133(2):701–764.

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/*

Downes, Alexander B. and Mary Lauren Lilley. 2010. "Overt Peace, Covert War?: Covert Intervention and the Democratic Peace." *Security Studies* 19(2):266–306.

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*

Escribà-Folch, Abel and Joseph Wright. 2010. "Dealing with tyranny: International sanctions and the survival of authoritarian rulers." *International Studies Quarterly* 54(2):335–359.

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.

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).".

Fuhrmann, Matthew and Yonatan Lupu. 2016. "Do Arms Control Treaties Work? Assessing the Effectiveness of the Nuclear Nonproliferation Treaty." *International Studies Quarterly* 60:530–39.

Gibler, Douglas M, Steven V Miller and Erin K Little. 2016. "An analysis of the militarized interstate dispute (MID) dataset, 1816–2001." *International Studies Quarterly* 60(4):719–730.

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* .

Kreps, Sarah E. and Matthew Fuhrmann. 2011. "Attacking the Atom: Does Bombing Nuclear Facilities Affect Proliferation?" *Journal of Strategic Studies* 34:161–187.

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

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.

Marinov, Nikolay. 2005. "Do economic sanctions destabilize country leaders?" *American Journal of Political Science* 49(3):564–576.

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.

McManus, Roseanne W. 2017. *Statements of resolve: Achieving coercive credibility in international conflict*. Cambridge University Press.

McManus, Roseanne W. 2018. "Making It Personal: The Role of Leader-Specific Signals in Extended Deterrence." *The Journal of Politics* pp. 000–000.

McManus, Roseanne W and Keren Yarhi-Milo. 2017. "The logic of "offstage" signaling: Domestic politics, regime type, and major power-protégé relations." *International Organization* 71(4):701–733.

Morgan, T Clifton, Navin Bapat and Yoshiharu Kobayashi. 2014. "Threat and imposition of economic sanctions 1945–2005: Updating the TIES dataset." *Conflict Management and Peace Science* 31(5):541–558.

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

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

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

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.

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

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

Rabe, Stephen G. 1988. "Eisenhower and Latin America : the foreign policy of anticommunism.".

Rauchhaus, Robert W. 2009. "Principal-Agent Problems in Humanitarian Intervention: Moral Hazards, Adverse Selection, and the Commitment Dilemma." *International Studies Quarterly* 53(4):871–884.

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

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

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

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

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." Australian Institute of International Affairs.

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

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

Times, New York. 1961. "U.S. PRO-CASTRO UNIT ASKS INQUIRY ON C.I.A.". January 6, 1961.
URL: *https://www.nytimes.com/1961/01/06/archives/us-procastro-unit-asks-inquiry-on-cia.html*

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

Warren, T Camber. 2014. "Not by the sword alone: Soft power, mass media, and the production of state sovereignty." *International organization* 68(1):111–141.

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.

7 Appendix

Outline:

7.1 Recap of model assumptions and results

For ease of reference, we begin by restating the setup and results of the formal model, then provide the proof for each in turn.

To recap, the sequence of moves is as follows:

1. The leader's type $\theta \in \{0, 1\}$, with $Pr(\theta = 1) = \pi$, is realized by Nature and observed privately by the leader.
2. The state variable $\omega \in \{0, 1\}$, with $Pr(\omega = 1) = \tau$ and the cost variable k_c, with $F^{\theta}(x) = Pr(k_c \leq k_c; \theta)$, are realized by Nature and observed privately by the leader.
3. The leader chooses whether to take public action $a_p \in \{0, 1\}$, which A observes, and covert action $a_c \in \{0, 1\}$, which A does not observe directly.
4. The policy outcome $y \in \{0, 1\}$ is realized, according to the probabilities given in (1).
5. The covert revelation $z \in \{0, 1\}$ is realized, with $Pr(z = 1|a_c) = a_c\lambda$.
6. The audience observes $(a_p, y, z) \in \{0, 1\}^3$, and chooses whether to punish or reward the leader, $r \in \{0, 1\}$
7. Payoffs are realized, with
$$U_A = \mathbb{1}[r = \theta]$$
and
$$U_L = y - a_c(k_c + \theta\delta) - a_p k_p + r\beta$$

Next, we restate Assumption 1 and discuss the justification for each parameter restriction.

Assumption 1 (Parameter restrictions) Throughout the analysis, assume the following:

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

(ii) $\alpha_0 < \min\left\{\alpha_c(1-\lambda), \alpha_p^1\alpha_c\right\}$

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

(iv) π is in an intermediate range:

– The lower bound of π is given by:

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

which implies that A's belief in the RE with no CS satisfies $\mu^{110} \geq \frac{1}{2}$ (see Lemma 6).

– The upper bound of π is defined implicitly, in the proof of Proposition 1.

Point (i) ensures that an equilibrium exists in which the scrupulous leader takes the public action if and only if it is effective (that is, an equilibrium in which the scrupulous leader does not "pander"; see discussion in the next section). Point (ii) means that the probably of success due to exogenous factors is relatively small. When it is violated, the leader's plausible deniability problem becomes trivial, as the audience becomes too inclined to grant the leader the benefit of the doubt. Point (iii) clarifies the distinction between "effective" vs. "ineffective" public action: the lower bound on k_p implies that ineffective public action ($a_p = 1$ when $\omega = 0$) is not worthwhile for policy reasons alone (a point that is central to our definition of a "cover story", as explained below), while the upper bound implies that effective public action is worthwhile.[29] Finally, point (iv) implies that it is neither too easy nor too difficult for the leader to avoid punishment from the audience, meaning that the audience can meaningfully influence the leader's behavior.

The main text imposed Assumption 2 (restricting attention to RE). Here we relax that assumption, and instead imposing a less demanding Assumption 3, which rules out some equilibria by imposing restrictions on off-path beliefs. All results reported below are supported under either assumption, or both.

Assumption 3 (Off-path beliefs) *Assume A assigns belief $\mu = \pi$ to any off-path information set with $z = 0$, and $\mu = 0$ to any off-path information set with $z = 1$.*

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

We now restate each of the formal results from Section 2 of the main text (along with Proposition 3, which did not appear in the main text), before presenting the proofs for each.

Proposition 1 (RE Existence). A responsive equilibrium always exists.

Corollary 1 (RE Optimality). Among all equilibria, the RE yields the best policy payoff for both the scrupulous leader and the audience. If α_0 is low, the RE yields the best overall payoff for the scrupulous leader.

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

[29] Note that the assumption that $\alpha_0 < \alpha_p^1\alpha_c$, from point (ii), implies that $\alpha_p^1(1-\alpha_c) < \alpha_p^1 - \alpha_0$; we include both expressions in point (iii) to clarify that they correspond to the marginal policy benefits of effective public action on its own ($\alpha_p^1 - \alpha_0$), or in combination with covert action ($\alpha_{pc}^1 - \alpha_c = \alpha_p^1(1-\alpha_c)$).

Note that Proposition 3 is a weaker but more general claim than Proposition 2 from the main text: Proposition 3 is a statement of CSE existence (without restricting attention to RE); while Proposition 2 says that, for $\lambda < \lambda^*$, all RE are CSE.

Proposition 2 (Equilibrium regions within RE). There exist thresholds λ^* and λ^{**} such that, within any RE:

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

Corollary 2 (CSE comparative statics, paraphrased). The threshold λ^* is increasing in α_c and decreasing in k_p; and if α_0 is small, it is increasing in β.

Corollary 3 (Cover Stories and Scrutiny). Consider any RE in which a cover story is played with positive probability (that is, when $\lambda < \lambda^*$). The audience's interim beliefs about the leader's scrupulousness upon observing public action (but before observing the outcome, or any revelation of covert action) are strictly less favorable than their interim beliefs upon observing no public action.

7.2 Proofs of main text results, Section 2

Remark 2 (Notation)

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

To preview the structure of the proofs:

- Lemmas 1 and 2 establish basic properties of the leader's and audience's best-responses, respectively.
- Lemmas 3 and 4 provide a general characterization of when cover stories can be used in equilibrium.
- Lemma 5 characterizes the leader's strategy within an RE.
- Lemma 6 characterizes the audience's beliefs within an RE.
- Proposition 1 builds on these lemmas to show that an RE always exists.

- Lemma 7 demonstrates that there exists a threshold λ^{**} such that, within any RE, the leader takes covert action with positive probability if and only if $\lambda < \lambda^{**}$.
- Lemma 8 demonstrates that if λ is below a threshold λ', then any RE must be a CSE.
- Proposition 2 follows directly from Lemma 7 and Lemma 8.
- Proposition 3 extends Lemma 8, to show that in *any* equilibrium, a CSE exists if and only if λ is below a threshold.
- Corollary 1 compares the scrupulous leader's expected payoff in the RE vs. any other equilibria that may exist.
- Corollary 2 characterizes how the λ threshold from Proposition 3 varies as a function of parameters α_c, k_p, and β.
- Corollaries 3 and 4 establish additional results regarding players' beliefs and behaviors within the RE/CSE.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

■

Remark 3 *An RE requires* $\tilde{k}_p^0 \leq k_p \leq \tilde{k}_p^1$

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

Remark 4 *Comparing across thresholds defined in Lemma 1:*

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

Lemma 2 *In every equilibrium:*

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

$$r = \begin{cases} 0, & \mu^h < \frac{1}{2} \\ 1, & \mu^h > \frac{1}{2} \end{cases} \tag{5}$$

- *Upon observing the direct revelation of covert action ($z = 1$), the audience fully punishes the leader ($r = 0$).*

Proof of Lemma 2: Equation (5) follows directly from the audience's utility function, given in (3). The second point of the lemma follows from the fact that L^1 never takes covert action, so the observation of $z = 1$ implies that $\mu = 0$ (either on-path by Bayes' Rule, or off-path given Assumption 3). ■

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

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

Lemma 4 *In any equilibrium:*

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

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

For the second point: In any equilibrium with $k_p > \tilde{k}_p^1$, L^1 never plays $a_p = 1$ on the path of play, so we must have that $s = t = 0$.[30] So we have $\hat{k}_p^0(s = t = 0) \leq \alpha_p^0(1 - \alpha_c)$, which again is less than k_p. ■

Lemma 5 (Leader strategies within an RE) *In any RE, the following strategy profile is the leader's best-response to the audience's strategy (q, s, t, v):*

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

[30] If L^0 also never played $a_p = 1$, so $a_p = 1$ is off-path and thus $s = t = \beta$ by Assumption 3, then we would have $\tilde{k}_p^1(s = t = \beta) = \alpha_p^1 - \alpha_0 + \beta - \alpha_0 q - (1 - \alpha_0)v \geq \alpha_p^1 - \alpha_0 > k_p$, contradicting $k_p < \tilde{k}_p^1$.

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

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

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

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

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

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

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

Proof of Lemma 6: Generally, observe that

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

$$Pr(a_p, y, z|\theta) = \sum_{\omega} Pr(a_p, y, z|\theta, \omega)Pr(\omega)$$

$$Pr(a_p, y, z|\theta, \omega) = \sum_{a_c} Pr(y, z|a_p, a_c, \theta, \omega)Pr(a_p, a_c|\theta, \omega)$$

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

$$F(x)\frac{(1-\alpha_{pc}^1)(1-\lambda)}{1-\alpha_p^1} + (1-F(x)) \leq F(x)\frac{\alpha_{pc}^1(1-\lambda)}{\alpha_p^1} + (1-F(x))$$

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

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

$$\frac{\pi}{(1-\pi)} - 1 \geq F(\ddot{k}_c^1)\left[\frac{\alpha_{pc}^1}{\alpha_p^1}(1-\lambda) - 1\right] \tag{6}$$

- The lefthand side of this inequality is positive, given $\pi > \frac{1}{2}$. If the quantity in the square brackets is nonpositive, the inequality is satisfied. If that quantity is positive, that means that $\lambda < \frac{\alpha_c(1-\alpha_p^1)}{\alpha_{pc}^1}$.
- Turning to the expression for $\ddot{k}_c^1$, we can see that it is strictly increasing in s for any $\lambda < \frac{\alpha_c(1-\alpha_p^1)}{\alpha_{pc}}$. Thus if (6) is satisfied for $s = \beta$, then it is satisfied for any s.
- The lower bound for π given in Assumption 1 is equivalent to (6) with $s = t = \beta$ plugged into the expression for $\ddot{k}_c^1$.

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

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

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

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

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

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

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

which we can see is positive.[31] Then to support the RE we need $k_p \leq \tilde{k}_p^1(t = v = \beta, s = s^*(q))$. To

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

see that this is satisfied, observe:

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

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

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

Lemma 7 (RE under high transparency) *There exists a threshold λ^{**} such that, within any RE, the leader takes covert action with positive probability if and only if $\lambda < \lambda^{**}$.*

Proof of Lemma 7: Consider an RE in which the leader never uses covert action. In this equilibrium, the audience holds beliefs $\mu^{a_p, y, z=0} = \pi$ for all a_p, y, and plays $q = s = t = v = \beta$ (as per Lemma 6). Since the leader never uses covert action, we know that

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

which rearranges to

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

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

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

Lemma 8 (CSE threshold within RE) *Within any RE:*

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

where

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

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

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

---

the previous case.

Proof of Lemma 8: First, observe that the expression for μ^{010} denotes A's belief $Pr(\theta = 1|a_p = 0, y = 1, z = 0)$ in the RE with no CS. It is strictly decreasing in q. If $\mu^{010}(q = 0) < \frac{1}{2}$, then A's best response is $q = 0$; if $\mu^{010}(q = \beta) > \frac{1}{2}$, then A's best response is $q = \beta$; otherwise, the equilibrium requires that $q = \hat{q}$, which is the unique q that solves $\mu^{010} = \frac{1}{2}$.

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

Alternatively, suppose we have an RE with CS played with positive probability. L^1's strategy does not change relative to the RE with no CS, whereas L^0's strategy shifts some probability from $a = (0, 1)$ to $a = (1, 1)$, which makes q weakly increase relative to the RE with no CS. In the CS, we know $k_p = \hat{k}_p^0$, which rearranges to $\lambda = \bar{\lambda}(q')$ for some $q' \geq q^*$. Because $\bar{\lambda}$ is decreasing in q, we know that $\lambda = \bar{\lambda}(q')$ implies $\lambda \leq \bar{\lambda}(q^*)$. This gives us the contrapositive of the first bullet point in the lemma. ■

Proof of Proposition 3 (CSE Existence): Recall:

- CSE existence requires $k_p = \hat{k}_p^0$, which in turn implies $k_p \leq \tilde{k}_p^1$ (as per Lemma 4).
- Proposition 1 showed that the RE (with $\tilde{k}_p^0 < k_p < \tilde{k}_p^1$) always exists.
- Lemma 8 showed that the λ^* threshold in the RE is $\bar{\lambda}(q^*)$.

We will consider separate equilibrium cases of: $k_p < \tilde{k}_p^0$; $k_p = \tilde{k}_p^0$; $\tilde{k}_p^0 < k_p < \tilde{k}_p^1$; and $k_p = \tilde{k}_p^1$. We will show:

- The λ^* threshold, below which a CSE is supported, is highest in the equilibrium with $k_p < \tilde{k}_p^0$, if such an equilibrium exists.
- If an equilibrium with $k_p < \tilde{k}_p^0$ does not exist, then the highest λ^* threshold (across all possible equilibria) is $\bar{\lambda}(q^*)$, as per Lemma 8,

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

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

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

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

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

narrower range of λ relative to the RE.

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

Proof of Corollary 2 (CSE Comparative Statics): Recall from Proposition 3 that λ^* is either equal to λ' or $\bar{\lambda}(q^*)$. In the case that $\lambda^* = \lambda'$, the result follows simply from taking partial derivatives. In the case that $\lambda^* = \bar{\lambda}(q^*)$, the comparative statics must account for both the direct effects of the parameters on $\bar{\lambda}$ (holding fixed q^*), and any indirect effects via q^*:

- k_p has only a direct effect.
- The direct and indirect effects of α_c work in the same direction.
- The direct and indirect effects of β work in opposite directions; but if α_0 is sufficiently small, then $q^* = 0$ for any β, which shuts down the indirect effect.

■

Proof of Corollary 1 (RE Optimality): The claim that the RE is yields the optimal policy payoff for the scrupulous leader follows trivially from the definition of the RE (the equilibrium in which the scrupulous leader plays $a_p = \omega$). The general strategy for proving the rest of the corollary (regarding overall payoff for the scrupulous leader) will be as follows:

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

We will consider two cases of the RE: first, when $\lambda \geq \bar{\lambda}(q^*)$, in which case L^0 never uses a cover story;[32] of and second, when $\lambda < \bar{\lambda}(q^*)$, in which case she uses a CS with positive probability.

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

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

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

- In an equilibrium with $k_p > \tilde{k}_p^1$, we have $s = t = 0$ (because the scrupulous leader never takes public action); so this is dominated by RE/CSE.
- In an equilibrium with $k_p \leq \tilde{k}_p^0$, we must have $v < \beta$ (because if $v = \beta$ then $\tilde{k}_p^0$ would be strictly less than k_p); in order to yield a higher payoff than the RE/CSE, we would need $s > s^*$; but given $v < \beta$, any s that would satisfy $k_p \geq \hat{k}_p^0$ would have to be lower than s^* (and any $s \geq s^*$ would yield $\hat{k}_p^0 > k_p$, which we established above cannot hold in equilibrium).
- In the best-case equilibrium with $k_p = \tilde{k}_p^1$, we have $\hat{k}_p^0 < k_p$,[33] and $t = v = \beta$,[34] and

$$s = \tilde{s} := \frac{k_p - (\alpha_p^1 - \alpha_0)(1-\beta)}{\alpha_p^1}$$

  Plugging in the upper bound for k_p and β from Assumption 1 into this expression, we find that $\tilde{s} < s^*$, which implies a lower expected payoff for L^1 than in the RE/CSE.

■

Proof of Corrollary 3 (Cover Stories and Scrutiny): The audience's interim beliefs upon observing $a_p = 1$, but before observing y or z, are given by

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

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

7.3 Proofs of main text results, Section 4

Corollary 4 (Correlation between public and covert action). Within an RE:

- If transparency is sufficiently high, covert action and public action are negatively correlated.
- If transparency is sufficiently low, covert action and public action may be positively correlated.

Proof of Corrollary 4 (Correlation between public and covert action): To show that (within an RE) covert action and public action are negatively correlated whenever λ is sufficiently high, we will start from the condition that $\lambda > \bar{\lambda}(q^*)$, which means a CS is never used. In this case, we want to show:

$$\begin{aligned} E[a_p | a_c = 1] &< E[a_p | a_c = 0] \\ \frac{Pr(a_p = 1, a_c = 1)}{Pr(a_c = 1)} &< \frac{Pr(a_p = 1, a_c = 0)}{Pr(a_c = 0)} \\ \frac{(1-\pi)\tau F(\ddot{k}_c^1)}{(1-\pi)[\tau F(\ddot{k}_c^1) + (1-\tau)F(k_c^*)]} &< \frac{\tau[\pi + (1-\pi)(1 - F(\ddot{k}_c^1))]}{\pi + (1-\pi)[\tau(1 - F(\ddot{k}_c^1)) + (1-\tau)(1 - F(k_c^*))]} \end{aligned}$$

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

[34] $t = v = \beta$ is pinned down by A's incentive-compatibility constraint, given that her beliefs in the corresponding information sets, μ^{100} and μ^{000}, are greater than $\frac{1}{2}$.

This rearranges to $F(\ddot{k}_c^1) < F(k_c^*)$, or

$$\alpha_c(1-\alpha_p^1) - \beta\lambda < \alpha_c - \alpha_0 + q[\alpha_c(1-\lambda) - \alpha_0] + \beta[(1-\alpha_c)(1-\lambda) - (1-\alpha_0)]$$

which is satisfied if

$$\lambda > \tilde{\lambda} := 1 + \left[\frac{-\beta\alpha_0 + \alpha_0 - \alpha_c\alpha_p^1}{\beta\alpha_c}\right]$$

where $\tilde{\lambda} < \frac{\alpha_c - \alpha_0}{\alpha_c}$, satisfying Assumption 1 (ii). Thus we can say that if $\lambda > \max\left\{\bar{\lambda}(q^*), \tilde{\lambda}\right\}$, then in an RE, covert action and public action are negatively correlated.

Next, we want to show that if λ is sufficiently low, there exist conditions under which covert action and public action are positively correlated. We will start from the condition that $\lambda < \bar{\lambda}(q^*)$, meaning CS is played with positive probability, and analogously to the previous case, we want to show

$$\begin{aligned}
E[a_p|a_c = 1] &> E[a_p|a_c = 0] \\
\frac{Pr(a_p = 1, a_c = 1)}{Pr(a_c = 1)} &> \frac{Pr(a_p = 1, a_c = 0)}{Pr(a_c = 0)} \\
\frac{(1-\pi)[\tau F(\ddot{k}_c^1) + (1-\tau)F(k_c^*)\hat{\sigma}]}{(1-\pi)[\tau F(\ddot{k}_c^1) + (1-\tau)F(k_c^*)]} &> \frac{\tau[\pi + (1-\pi)(1-F(\ddot{k}_c^1))]}{\pi + (1-\pi)[\tau(1-F(\ddot{k}_c^1)) + (1-\tau)(1-F(k_c^*))]}
\end{aligned}$$

This ultimately rearranges to

$$Pr(a_p = 1|a_c = 1, \omega = 0, \theta = 0) = \hat{\sigma}_p > \frac{\tau(F(k_c^*) - F(\ddot{k}_c^1))}{F(k_c^*)[1 - (1-\pi)\tau F(\ddot{k}_c^1) - (1-\pi)(1-\tau)F(k_c^*)]} \tag{7}$$

Numerical simulations demonstrate that there exists a range of parameter values for which this condition is satisfied. For instance, consider the following:

- Let $\beta = 0.5$, $\pi = 0.55$, $k_p = 0.25$, $\tau = 0.5$, $\alpha_0 = 0.05$, $\alpha_c = 0.5$, $\alpha_p^0 = 0.2$, $\alpha_p^1 = 0.6$, $\lambda < 0.9$. Let F be a uniform distribution on [0,0.7]. These values satisfy parts (i), (ii), and (iii) of Assumption 1, and they satisfy the lower bound of π in part (iv).
- Consider $\lambda = 0$. We can see that $0 < \hat{k}_p^0(s = t = v = \beta, q = 0)$, so the RE requires that a CS be played with positive probability.
- A CSE can be supported in which A plays $t = v = \beta$, $q = 0$, and $s = s^* = \frac{1}{3}$, and L plays $\hat{\sigma}_p = \frac{4}{9} \approx 0.444$. We can confirm that these values satisfy A's incentive-compatibility conditions (yielding $\mu^{100} > \mu^{110} = \frac{1}{2} > \mu^{010}$, thus satisfying the implicit upper bound on π from Assumption 1 (iv)), and L's indifference condition (yielding $k_p = \hat{k}_p^0(s = s^*, q = 0, t = v = \beta)$).
- Plugging in values to the righthand side of (7) gives ≈ 0.151, which is less than the equilibrium value of $\hat{\sigma}_p$.

■

Restating Remark 1 from Section 4:

- The observed difference in mean success rates (ODIM) can be ether positively or negatively biased for the true ATT of overt action, with the magnitude and direction of bias varying with the level of transparency.

Here we derive the expressions used to compute the Observed Difference in Means (ODIM) and the Average Treatment Effect on the Treated (ATT), as reported in Figure 4.

Let ΔY denote the unit-level treatment effect of overt action, which we define as: the counterfactual difference between the outcome that would occur if overt action were taken, vs. if overt action were not taken, holding fixed the state ω and whether or not covert action was taken.[35] Let $E[\Delta Y|a_p = 1]$ denote the ATT of overt action.

When $\lambda > \lambda^{**}$: covert action is never used, and overt action is only used when $\omega = 1$, so ODIM = ATT = $\alpha_p^1 - \alpha_0$

When $\lambda^* < \lambda < \lambda^{**}$:

$$\begin{aligned} ATT &= E[\Delta Y|a_p = 1] \\ &= E[\Delta Y|a_p = 1, a_c = 1]Pr(a_c = 1|a_p = 1) + E[\Delta Y|a_p = 1, a_c = 0]Pr(a_c = 0|a_p = 1) \\ &= \alpha_p^1(1 - \alpha_c)F(\ddot{k}_c^1) + (\alpha_p^1 - \alpha_0)[1 - F(\ddot{k}_c^1)] \end{aligned}$$

$$\begin{aligned} ODIM &= E[y|a_p = 1] - E[y|a_p = 0] \\ &= \sum_{a_c} \Big(E[y|a_p = 1, a_c]Pr(a_c|a_p = 1) - E[y|a_p = 0, a_c]Pr(a_c|a_p = 0) \Big) \\ &= \alpha_{pc}^1 F(\ddot{k}_c^1) + \alpha_p^1[1 - F(\ddot{k}_c^1)] - \alpha_c F(k_c^*) - \alpha_0[1 - F(k_c^*)] \end{aligned}$$

When $\lambda < \lambda^*$:

$$ATT = E[\Delta Y|a_p = 1, \omega = 0]Pr(\omega = 0|a_p = 1) + E[\Delta Y|a_p = 1, \omega = 1]Pr(\omega = 1|a_p = 1)$$

$$Pr(\omega = 1|a_p = 1) = \frac{\tau}{\tau + \hat{\sigma}_p F(k_c^*)(1 - \tau)}$$

$$E[\Delta Y|a_p = 1] = \frac{\alpha_p^0(1 - \alpha_c)\hat{\sigma}_p F(k_c^*)(1 - \tau) + \tau \Big(F(\ddot{k}_c^1)\alpha_p^1(1 - \alpha_c) + (1 - F(\ddot{k}_c^1))(\alpha_p^1 - \alpha_0) \Big)}{\tau + (1 - \tau)\hat{\sigma}_p F(k_c^*)}$$

[35] Insofar as we think of the leader's choice to use covert action as being causally downstream from the choice to use overt action, it might make sense to instead characterize this treatment effect as a Controlled Direct Effect; see Acharya, Blackwell and Sen (2016).

$$
\begin{aligned}
E[y|a_p] &= \sum_{\omega} E[y|a_p,\omega]Pr(\omega|a_p) \\
&= \frac{1}{Pr(a_p)}\sum_{\omega} E[y|a_p,\omega]Pr(a_p|\omega)Pr(\omega) \\
E[y|a_p=1] &= \frac{E[y|a_p=1,\omega=1]\tau + E[y|a_p=1,\omega=0](1-\tau)F(k_c^*)\hat{\sigma}_p}{\tau+(1-\tau)F(k_c^*)\hat{\sigma}_p} \\
&= \frac{[F(\ddot{k}_c^1)\alpha_{pc}^1 + (1-F(\ddot{k}_c^1))\alpha_p^1]\tau + \alpha_{pc}^0(1-\tau)F(k_c^*)\hat{\sigma}_p}{\tau+(1-\tau)F(k_c^*)\hat{\sigma}_p} \\
Pr(a_c|a_p=0,\omega=0) &= \frac{Pr(a_c,a_p=0|\omega=0)}{Pr(a_p=0|\omega=0)} = \frac{Pr(a_c,a_p=0|\omega=0)}{1-F(k_c^*)\hat{\sigma}_p} \\
E[y|a_p=0] &= E[y|a_p=0,\omega=0] \\
&= \sum_{a_c} E[y|a_c,a_p=0,\omega=0]Pr(a_c|a_p=0,\omega=0) \\
&= \frac{\sum_{a_c} E[y|a_c,a_p=0,\omega=0]Pr(a_c,a_p=0|\omega=0)}{1-F(k_c^*)\hat{\sigma}_p} \\
&= \frac{\alpha_c F(k_c^*)(1-\hat{\sigma}_p)+\alpha_0[1-F(k_c^*)]}{1-F(k_c^*)\hat{\sigma}_p} \\
ODIM &= E[y|a_p=1] - E[y|a_p=0] \\
&= \frac{[F(\ddot{k}_c^1)\alpha_{pc}^1 + (1-F(\ddot{k}_c^1))\alpha_p^1]\tau + \alpha_{pc}^0(1-\tau)F(k_c^*)\hat{\sigma}_p}{\tau+(1-\tau)F(k_c^*)\hat{\sigma}_p} - \frac{\alpha_c F(k_c^*)(1-\hat{\sigma}_p)+\alpha_0[1-F(k_c^*)]}{1-F(k_c^*)\hat{\sigma}_p}
\end{aligned}
$$

7.4 Pure Moral Hazard

Here we consider an alternative model setup, in which the agency problem is one of pure moral hazard, rather than adverse selection. The model setup is as follows:

- The game sequence follows Figure 1, with the exception that the leader is commonly known to be of type $\theta = 0$, i.e. all leaders are unscrupulous.
- Leader payoff is still given by (4).
- Audience payoff is: $U_A = -a_c$; that is, the audience simply seeks to minimize the leader's use of covert action.

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

Notation. Let

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

and let

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

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

Assumption 4 (MH Parameter Restrictions)

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

(ii) $\alpha_0 < \min\left\{\alpha_c(1-\lambda), \alpha_p^1\alpha_c\right\}$

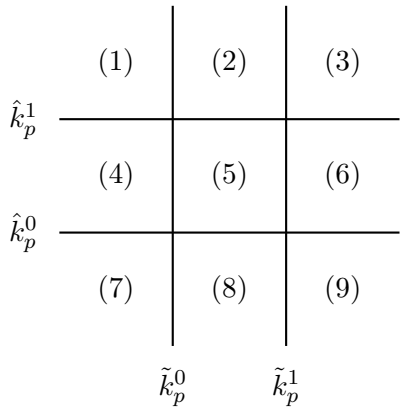
(iii) k_p *is in an intermediate range,* $\alpha_p^0 < k_p < \min\left\{\alpha_p^1 - \alpha_0, \alpha_p^1(1-\alpha_c)\right\}$

(iv) $\underline{k_c}$ *is in an intermediate range,* $\alpha_c(1-\alpha_p^1) \leq \underline{k_c} \leq \min\left\{k_c^\dagger, k_c'\right\}$

The first three points are the same as in Assumption 1. Intuitively, the fourth point implies two things: (i) there exists an audience strategy which can fully disincentivize the leader from taking covert action when public action is effective ($\omega = 1$); and (ii) there does not exist an audience strategy which can fully disincentivize the leader from taking covert action when public action is ineffective ($\omega = 0$).

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

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



Lemma 9 (CSE under pure moral hazard) *If* $\lambda < \bar{\lambda}(q=0)$*, as defined in Lemma 8, then there exists a CSE which yields* $Pr(a_c) \leq (1-\tau)F(k_c^\dagger)$.

Proof of Lemma 9: Propositions 1 and 3 demonstrated that when $\lambda < \bar{\lambda}(q=0)$, a CSE exists, characterized by: $t = v = \beta$, $q = 0$, and $s \in (0, \beta)$ that satisfies $k_p = \hat{k}_p^0 < \hat{k}_p^1$ and $\tilde{k}_p^0 \leq k_p \leq \tilde{k}_p^1$ (corresponding to case (5) in Figure 5). This was proven in the adverse selection setting, in which the audience's equilibrium strategy was pinned down by sequential rationality (i.e. maximizing (3)) given their beliefs about L's type. In the present setting, any audience strategy is permissible in equilibrium (including, of course, the audience strategy posited in Propositions 1 and 3). It is straightforward to see that the specified CSE is still supported with the additional restriction on $\underline{k_c}$ asserted in Assumption 4 (iv).

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

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

Proof of Proposition 4: Following Lemma 9, all that remains to show is that when $\lambda \leq \bar{\lambda}(q=0)$, there does not exist any non-CSE equilibrium with $Pr(a_c) < (1-\tau)F(k_c^\dagger)$. We can consider each case depicted in Figure 5:

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

This exhausts all cases. ■