Appendix

Table of Contents

A Formal appendix

A.1 Characterizing the distributions

In the manuscript we focus on three quantities: e_0, e_1, λ. Here we provide more information about their properties, which follow from the definition of Bayes' Rule. We will use these properties in constructing some of the proofs.

Remark on $p(\pi)$. The prior is given as $p(\pi)$ with an expected value, $e_0 = \mathbb{E}[\pi] = \int \pi p(\pi) d\pi$.

Defining the posterior We notate the posterior distribution given an observed m, $p_1(\pi|m) \propto f_{m|\pi}(m|\pi)p(\pi)$, suppressing the proportionality constant. The post-message expected value of π is, $\mathbb{E}[\pi|m] \propto \int \pi p_1(\pi|m) d\pi$.

Remark Properties of posterior belief. $\partial \mathbb{E}[\pi|m]/\partial m > 0$, and that $e_0 > \mathbb{E}[\pi|m]$ if $m < e_0$, $e_0 < \mathbb{E}[\pi|m]$ if $m > e_0$, $e_0 = \mathbb{E}[\pi|m]$ if $m = e_0$.

Define m^* as the message that satisfies $\mathbb{E}[\pi|m = m^*] = c_D - \theta$. Note, that if m^* can be defined, it is unique, that every message $m > m^* \implies \mathbb{E}[\pi|m > m^*] > c_D - \theta$, and that m^* is increasing in c_D and decreasing in θ. Whether m^* can be defined depends on p, c_D, θ. If p is binomial, for example, m^* is not necessarily definable. But it can be defined in many important cases. For example, if p is normal, then m^* can always be defined.

Remark Properties of λ. If m^* exists, $\partial \lambda / \partial k_D < 0$.

Remark Properties of λ. If m^* exists, we can re-write $\lambda = \mathbb{E}[m \geq m^*]$. Then, $\partial m^*/\partial \lambda < 0$. If m^* does not exist, then $\lambda = 0$.

Remark Properties of e_1. e_1 is defined if m^* exists. If e_1 is defined, it must be a real-valued number that satisfies $e_1 > c_D - \theta$.

A.2 Lemma 1: When open institutions do not innovate

In the open institution, there are three strategy profiles that can lead to innovation. In the first R selects research, D does not approve research, then D selects innovation absent research. If this was on path, then D could not profitably deviate to rejecting innovation at the final decision-node. But D prefers to deviate to rejecting innovation if $e_0 - c_D < 0 \equiv e_0 < c_D$. This cannot be satisfied if condition 1 is.

In the second, R selects research, D approves research, then D approves innovation after research. Off the path, if D does not research D approves innovation. We cannot support the off-path action if condition 1 holds.

In the third, R selects research, D approves research, then D approves innovation after research. Off the path, if D does not research D does not innovate (condition 1 is satisfied). In this pathway, D approves innovation at the final on-path node if $\mathbb{E}[\pi|m]+\theta-c_D-k_D > -k_D \equiv \mathbb{E}[\pi|m] > c_D-\theta$. Note, we use this to define λ.

Working backwards, D's expected utility for authorizing research given expectations of on-path play is $\lambda(e_1+\theta-c_D)-k_D$. If instead, D does not research he gets 0. This solves for condition 2.

This completes the proof.

A.3 Proposition 1: Secrecy facilitates innovation (+ when it does not)

We re-state the equilibrium strategies. R's strategy is to select secret research. D's on-path strategy is to select secret research, and then approve innovation post-secret research if research yields a signal that shifts D's posterior belief $\mathbb{E}[\pi|m] > c_D-\theta$, and reject the innovation otherwise. Off the path, if R asks for permission to conduct research, D does not approve research and does not approve innovation.

Notice that if R does ask for permission, we are in a sub-game that exactly reflects the open institution. It follows that D's off-path strategy to reject research and reject innovation is supported if Conditions 1, 2 are. This yields a pay-off of $0, 0$. Similarly, if R rejects an idea at the first node, pay-offs are 0 for both players.

Turning to on-path actions, in the final node, D approve an innovation iff $\mathbb{E}[\pi|m]+\theta-c_D-k_Dx \geq -k_Dx$.

Working backwards, consider R's on-path decision to engage in secret research. R's value from secret research is: $(1-\lambda)0+\lambda(e_1+\theta-c_Rx)-k_R$. R prefers this to all her other options (which each yield 0) when equilibrium condition 3 is satisfied. It follows that R cannot profit from deviating to open research, or scrapping the project under the conditions stated in the equilibrium.

Later, we will need to know when we do not observe secret innovation in the secret institution.

Lemma 3 *If conditions 2 is violated, then we cannot support secret innovation in the secret institution.*

When condition 2 is violated, we cannot support an SPE where D rejects open research. Thus, in every SPE D will approve open research if R selects open research. In this case, R's expected utility from requesting open research is $\lambda(e_1+\theta-c_Rx)-k_Rx$, which strictly dominates R's value from secret research.

Lemma 4 *If condition 2 is violated and condition 1 is not, and*

$$\lambda > \frac{k_rx}{e_1+\theta-c_Rx} \tag{6}$$

then then we can support open research in the secret institution. In equilibrium, D approves open research, does not approve innovation without research, and approves innovation following research (secret or open) iff $\mathbb{E}[\pi|m] > c_D-\theta$. R selects open research.

We've already shown that: (i) when condition 1 is violated, D rejects innovation absent research; and that (ii) when condition 2 is violated, D prefers to approve open research than reject it given (i); and that (iii) R strictly prefers open research given to secret research given (ii). Finally, R prefers open research to scrapping the project if $\lambda(e_1 + \theta - c_R x) - k_R x > 0$, which solves for the equilibrium condition. This completes the proof.

A.3.1 Existence of Proposition 1

We have solved for a set of conditions where Proposition 1 is incentive compatible. We now verify that these conditions can be satisfied for some parameters. To be clear, these are not exhaustive conditions. Our only goal is to demonstrate that the equilibrium conditions can be satisfied.

Remark Assume a prior distribution $\pi \sim \mathcal{N}(0, \sigma_0)$, and parameters $\theta > c_D > c_R x$. Then, there must exist a $k_R, k_D > 0$ for which equilibrium conditions (1), (2) and (3) are simultaneously satisfied.

First note that $e_0 = 0 \implies e_0 < k_D$. Thus, condition (1) is satisfied.

We now prove that m^* must exist given the assumed parameters and distributions and discuss its implications for e_1, λ. Note the posterior distribution is

$$\mathcal{N}(\frac{\sigma m}{\sigma + \sigma_0}, \frac{1}{\sigma + \sigma_0})$$

with an expected value of $\frac{\sigma m}{\sigma + \sigma_0}$. Note because the domain of both prior and posterior cover all real valued numbers, we can always find an m that satisfies:

$$\frac{\sigma m}{\sigma + \sigma_0} = c_D - \theta \implies \exists\ m^*, m^* := \frac{(c_D - \theta)(\sigma + \sigma_0)}{\sigma}$$

This implies λ and e_1 exists and λ satisfies > 0. By construction, if e_1 exists, then $e_1 + \theta - c_D > 0 \implies e_1 + \theta - c_R x > 0$.

We now turn to conditions (2) and (3), which we re-write as:

$$\lambda(e_1 + \theta - c_D) < k_D$$

$$k_R < \lambda(e_1 + \theta - c_R x)$$

The only restriction on k_D, k_R is that they must be positive. We can trivially find a k_D sufficiently large to satisfy condition (2). To satisfy (3) a k_R must exist that satisfies $k_R \in (0, \lambda(e_1 + \theta - c_R x)$. We've shown that $\lambda > 0, e_1 + \theta - c_R x > 0 \implies \lambda(e_1 + \theta - c_R x) > 0$. Thus, the open interval $(0, \lambda(e_1 + \theta - c_R x)$ is always defined under the stated conditions.

A.4 Two pathways to innovation

In this section we explain how we derive our empirical implications from the main model. The basic idea is to conjecture a set of parameters where secrecy facilitates innovation (i.e, conditions 1, 2, and 3 all hold), then show how taking certain parameters to their limits implies we must violate either conditions 1, 2, and 3. If we violate them, then secrecy cannot faciliate innovation.

The first pathway describes different features of $p()$. Our first bullet point relates to e_0, the expected value of p. Define a second distribution p_α as identical to p in functional form, but with a shifted mean $\alpha \in \mathcal{R}$. That is, for an arbitrary input a, $p(a) = p_\alpha(a - \alpha)$. Note the prior expected value of $p_\alpha = e_0 + \alpha$.

Define the standard deviation of $p()$ as σ_p, not that it equals the standard deviation of p_α.

We can re-write the first bullet point in pathway 1 as follows. Suppose $\alpha = 0$, and otherwise we take a list of parameters that meet the conditions for secret innovation outlined in proposition 1. The claim in the bullet is that we will violate at least one equilibrium condition if $\alpha \to \infty, -\infty$.

Starting with the upper bound, there exists a $\bar{\alpha}$ large enough so that for any $\alpha > \bar{\alpha}$ inequality of condition 1 is violated. Since it is violated, we cannot say that innovation does not occur in the open institution, and therefore we cannot say that secrecy facilitates innovation.

Turning to the lower bond, there exists a $\underline{\alpha}$ low enough so that for any $\alpha < \underline{\alpha}$ inequality of condition 3 is violated. Both λ, e_1 are a function of α. By Bayes' Rule, as $\alpha \to -\infty$, $\lambda \to 0$. As stated above, as $\alpha \to -\infty$ e_1 is either undefined (which assures condition 3 cannot be satisfied) or must satisfy $e_1 > c_D - \theta$, as desired.

Our second bullet point relates to the standard error of p, σ_p. Define a second distribution p_β as p with a resolution parameter $\beta > 0$. More precisely, for an arbitrary input a, $p(e_0 + a) = p_\beta(e_0 + \beta a)$. Notice that the expected value of p_β is equal to e_0. We also define the standard error as σ_β. For $\beta < 1, \sigma_\beta < \sigma$, and for $\beta > 1, \sigma_\beta > \sigma$.

We can re-write the second bullet point as follows. Suppose a list of parameters where condition 1,2 and 3 are satisfied and replace p with p_β, $\beta = 1$. There exists a $\underline{\beta}$ small enough so that for any $\beta < \underline{\beta}$ that violates condition 3.

Taking $\beta \to 0 \implies \mathbb{E}[\mathbb{E}[\pi|m]|p_\beta] \to e_0$. Thus, taking $\beta \to 0$, if $e_0 < c_D - \theta \implies \lambda \to 0$ and e_1 is undefined. Note pathway 2 caveats that as θ cannot be too large, and we can more precisely characterize that as the condition $\theta < c_D - e_0$. Substantively, this means that the improvement value of research is insufficient to induce D to approve a program. This is consistent with the overall message of the argument, wherein research reveals information.

The second pathway simply highlights that proposition 1 can hold when research is costly and the idea is promising. It starts with the premise that managers know a project shows promise once it is improved by research ($e_0 + \theta >> 0$). However, they know that the research involves political costs that outweigh the amount that research improves the project (i.e. $\theta \leq k_D$).

We start by noting that that for any initial expectation of success e_0, and amount that research improvement θ, there exists a sensitivity to the costs of authorization c_D for which condition 2 holds. Similarly, for any e_0, there exists a sensitivity to the costs of research $k_D > e_0$ for which condition 1 holds.

Now focusing only on conditions where $e_0 > 0$ and condition 2 holds, and where m^* exists. Then there exists a $k_R \to 0$ for which we can satisfy 3. The proof follows instantly from the proof of proposition 1 (especially noting that 3 is always satisfied if $k_R = 0$.). So long as there is some chance that research will convince D, we can find a researcher sufficiently insensitive to costs who is willing to research given that small chance.

A.5 Expectation 1 and 2

Expectation 1 is validated if the λ values that support open research are larger than the λ values that support secret research. Expectation 1 follows from three facts. (1) Secret research cannot occur if condition 2 holds, which places an upper bound on λ. (2) Open research only requires that λ is sufficiently large (lemma 4). (3) $\lambda = Pr.$(Research Approved).

Turning to expectation 2. To provide some intuition, Expectation 2 takes the perspective of an outside observer who knows the true value of π. Holding the value of π constant at an effect that is sufficiently effective (which we shall define below), we ask the outside observer to considers two worlds: one in which secret research appears on path (call it the counter-factual case), and another where open research appears on path (call it the baseline case). In both cases innovation can occur with probability, but only if the message is sufficiently strong. How confident is the outside observer that innovation will actually occur in each case given that π is known? Then, if we increase π even more, how does it affect the outside observer's confidence that innovation will occur in the baseline relative to the counterfactual? The basic idea is that the minimum message m that will induce innovation in the counterfactual world must be stronger than the minimum message necessary to induce innovation in the baseline. Given that $m \sim \mathcal{N}(\pi, \sigma)$, increasing π has a greater impact on the observer's expectation of innovation in the counter-factual case.

To make this claim more precise way, we first must narrow our focus to two comparable cases. One that leads to secret research on path (call it the counter-factual case) and the other that leads to open research on path (call it the baseline case). To do it, we again utilize distribution that are identical in their function forms, but vary in their expectations. We call p our baseline distribution. We assume it is supported positively on $\mathbb{R}$. We then define a counter-factual distribution p_α as identical to p in functional form, but with a shifted mean $\alpha < 0$. That is, for an arbitrary input a, $p(a) = p_\alpha(a - \alpha)$. To differentiate between cases, we index expectation derived from p_α as $\lambda_\alpha, e_{0\alpha}, e_{1\alpha}$. Given $\alpha < 0$, note that $e_{0,\alpha} < e_0, e_{1,\alpha} < e_1, \lambda_\alpha < \lambda$.

We assume fixed values of prior parameters $k_i, c_i, x, \theta, \sigma$, and the functional form of p. We then assume that given the baseline case of p we satisfy the conditions for open research characterized in lemma 4, and that given the counterfactual case of p_α we can we satisfy the conditions for secret research characterized in proposition 1.

In both cases, research appears on path and there is a positive probability that D approves innovation occurs post-research. We've shown that in either case, D approves innovation post-research iff: $\mathbb{E}[\pi|m, p] > c_D - \theta$, where m is a function of π. Define $m^\dagger$ as the message necessary to satisfy $\mathbb{E}[\pi|m^\dagger, p] = c_D - \theta$, and $m^\dagger_\alpha$ as the message necessary to satisfy $\mathbb{E}[\pi|m^\dagger_\alpha, p_\alpha] = c_D - \theta$. By construction of the counter-factual case, $m^\dagger_\alpha > m^\dagger$.

Suppose an outside observer knows the true $\pi = \pi_x$. That outside observer's pre-research expectation that innovation occurs in our two worlds is characterized as:

$$\omega_x = \int_{m^\dagger}^{\infty} \mathcal{N}(\pi_x, \sigma) dm$$

$$\omega_{x,\alpha} = \int_{m_\alpha^\dagger}^{\infty} \mathcal{N}(\pi_x, \sigma) dm$$

These are the beliefs that a message $m > m^\dagger$ or $m > m_\alpha^\dagger$ will occur that will give D enough confidence to approve innovation. We similarly $\omega_y, \omega_{y\alpha}$ for $\pi_y > \pi_x$.

We can re-state expectation 2 as follows: Contrasting our two worlds, so long as the true effect of innovation is sufficiently large ($\pi_x > m^\dagger$), then increasing the true effect of innovation from $\pi = \pi_x \rightarrow \pi_y$ raises the probability of innovation in the counterfactual world more than the baseline world:

$$\omega_y - \omega_x < \omega_{y\alpha} - \omega_{x\alpha}$$

Re-arranging this term, the claim is true if:

$$\int_{m^\dagger}^{m_\alpha^\dagger} \mathcal{N}(\pi_y, \sigma) dm > \int_{m^\dagger}^{m_\alpha^\dagger} \mathcal{N}(\pi_x, \sigma) dm$$

true if $\pi_x > m^\dagger$. As desired.

A.6 Proposition 2: Monitoring and principal-agent dynamics

To start, we re-state the timing and information of the model more precisely.

Initially, Nature draws two random variables:

- $\pi \sim p()$ (unobserved by R or D)
- $k \sim f()$ (privately observed by R).

Then the first decision node is R's, where R decides between open research, secret research, or scrapping the project.

If R pursues open research:

- D observes R's request, and the value of k,
- the game proceeds as in baseline starting at the node where D can approve open research/not.

If R chooses to scrap the project,

- D observes that R has not conducted open research, but not k.

- D can monitor R or not. If D monitors, R's choice (scrap project) and k is observed by D and not observed otherwise.
- Regardless of D's choice to monitor the game ends with payoffs $0, 0$.

If R chooses secret research,

- D observes that R has not conducted open research, but not k.
- D can monitor R or not. If D does not monitor,
  - Neither R's choice (secret research) nor k are observed.
  - the game proceeds with secret research as in the baseline model with Nature's draw $m|\pi$. All payoffs are identical to the baseline.
- If D monitors:
  - D observes both R's choice of secret research, and k. D can chose to shut down research or not.
  - If D shuts down research
    - R incurs k, D incurs no research cost.
    - No player observes m.
    - D is given the choice to approve innovation or not, given $\mathbb{E}[\pi] = e_0$.
  - If D does not shut down research.
    - R incurs xk, D incurs k
    - Both players observe m
    - D is given the choice to approve innovation or not, with $\mathbb{E}[\pi|m]$.

Once research has occurred (or if it does not), the incentives for choices are equivalent. Thus, D will not approve innovation absent research if condition 1 holds. D will only approve innovation post research if $\mathbb{E}[\pi|m] > \theta - c_D$.

We now turn to D's decision to approve research. There are two ways D has the option to approve research. First, R may select open research. In this case, D will only approve if

$$\lambda[e_1 + \theta - c_D] > k \tag{7}$$

This is a re-statement of condition 2 subbing $k_D = k$. Second, R selects secret research and D monitors. In this case, D also only approves research if 7 holds.

We now conjecture that D will not monitor and identify R's incentive to select secret research, open research, or scrap the project for different values of k, assuming the conjecture holds. In a moment, we will consider D's incentive to monitor. We've shown if condition 7 is violated, then R's expected value from asking for open research is 0. R's value for scrapping a project is also 0. Thus for a k that violates condition 7, we could support either choice if 0 is better than R's expected utility from secret research. We've show above that if D will approve open research (7 holds), then R strictly prefers open research to secret research. Given these results, we need only consider when

R prefers secret research to a payoff of 0. R prefers secret research to if $\lambda[e_1 + \theta - c_R x] > k$. We assumed, $0 < \lambda[e_1 + \theta - c_D] < \lambda[e_1 + \theta - c_R x]$, which imposes an order on these conditions.

Putting it altogether, if D will not monitor, we can support R's on-path strategy, defined by two cut points on k. Let $\underline{k} = \lambda[e_1 + \theta - c_D]$. If $k < \underline{k}$, R selects open research, and D approves, as desired. Let $\bar{k} = \lambda[e_1 + \theta - c_R x]$. If $k > \bar{k}$, D will not approve open research if asked, R is indifferent between asking and being rejected and open research, and both these options are better than secret research. In equilibrium, we conjecture that R scraps the idea. If $k \in [\underline{k}, \bar{k}]$, R pursues secret research.

We now resolve our conjecture that D will not monitor in the case that D has not observed research. Off path, if D monitors after failing to observe open research, D gets 0 for any potential value of k. If $k \in [\underline{k}, \bar{k}]$, D rejects, leading to 0. If $k > \bar{k}$, R did not research and the idea is scrapped. D's expected utility from on path play (not monitoring) is:

$$pr[k > \bar{k}|nor] \times 0 + pr[k \in [\underline{k}, \bar{k}]|nor](\lambda(e_1 + \theta - c_D) - x\mathbb{E}[k|sr, nor])$$

Here $pr[k \in [\underline{k}, \bar{k}]|nor]$ is D's expectation R undertook secret research given that D did not observer research (nor represents no research observed). Then $\mathbb{E}[k|sr]$ is D's expected value of k given the values of k for which R conducted secret research under the condition that no research was observed (sr represents secret research occured). D prefers not monitoring to monitoring given D did not observe research if:

$$\frac{\lambda(e_1 + \theta - c_D)}{x} > \mathbb{E}[k|sr, nor] \tag{8}$$

as stated in the equilibrium.

A.7 Lemma 2: Researcher can fabricate her report

First we more fully specify the set-up of the extension with some reference to the baseline presented in Figure 1.

- D can costlessly set k_R, c_R (which represents a manager hiring a particular researcher).
- $\pi \sim p()$ (unobserved by R or D)
- R selects between open research, secret research, or scrapping the project.
  - Open research and scrapping the project proceed identically as in the baseline presented in Figure 1.
- If R selects secret research, R privately observes $m \sim \mathcal{N}(\pi, \sigma)$
- R writes costless message $m_R \in \mathcal{R}$, which is public.
- D decides to innovate or not.

The payoffs in this extension are identical to the baseline model, with the subtle difference that k_R, c_R are endogenous to D's choice.

The model includes a new information structure such that R has the same information as in the baseline model. But R also has private information about the message m in the case of secret research.

The new information structure means we must provide additional information about beliefs. We continue to define $e_0 = \mathbb{E}[\pi|p]$, $\lambda = pr(\mathbb{E}[\pi|m] > c_D - \theta)$, $e_1 = \mathbb{E}[\mathbb{E}[\pi|m]|\mathbb{E}[\pi|m] > c_D - \theta]$.

These expectations will play the same role in the analysis, in the event of open research, and for R's choice to engage in secret research. However, beliefs could deviate following secret research. We define $e_R = \mathbb{E}[\pi|m]$, as R's expected value of π given R has observed research. We define, $e_D = \mathbb{E}[\pi|m_R, s^R]$ as D's expected value of π given D observes R's message and R's strategy.

We define an honest researcher as one who sends the message $m_R = m \forall m$. We define a trust-worthy researcher as one that induces $e_D = e_R|m_R, s^R$ for all possible m. Meaning that D's beliefs match R's beliefs at the moment D must chose to approve innovation or not.

Lemma 5 *If conditions 1-3 hold, for $c_R = c_D/x$, and*

$$\lambda > \frac{k_D x}{e_1 + \theta - c_D} \tag{9}$$

then the following strategies are supported as a PBE.

D sets $k_R < \lambda(e_1 + \theta - c_R x), c_R = c_D/x$, then

- *D approves innovation following secret research iff $m_R \geq m^*$. Off-path, D rejects open research, and rejects innovation absent research.*
- *R selects secret research, and sets $m_R = m$.*

If, off-path, D sets $k_R > \lambda(e_1 + \theta - c_R x)$, or $c_R \neq c_D/x$, then we revert to the following off-path strategies:

- *R selects scrap the project, and sends a message $m_R \sim p(\pi)$ which is not conditioned on then observed m, and covers all feasible messages with positive probability (i.e, no off-path messages).*
- *D rejects research and innovation at every decision-node.*

Remark In equilibrium, R is honest $m_R = m$ and trustworthy ($e_R = e_D$).

We claim that if D deviates by setting the incorrect k_R, c_R, then R scraps the idea at the first decision node, leading to a payoff of 0 for both players. This is supported by a series of other off-path actions. We now solve for this off-path profile. Remaining in the case where D deviates to setting an off-path cost profile, we conjecture that if R did pursued secret research given an incorrect cost profile, that R also sends a babbling message that covers all feasible messages, and

D rejects innovation. Note because R's message is babbling, $e_D = e_0$. Thus, D rejects innovation if $e_0 < \theta - c_D$, true if conditions 1, 2 hold. If D will not approve innovation for any m, R strictly prefers to scrap the idea over secret innovation, as desired. We note that no matter R's cost profile, D rejects open research if condition 1 - 3 hold. Thus, R also cannot profitably deviate to open research, as desired. It follows that if D sets the incorrect cost profile, that we can support a strategy profile where players expect 0.

We now turn to the on-path case where D sets $k_R < \lambda(e_1 + \theta - c_R x), c_R = c_D/x$. I claim that $c_R = c_D/x$ induces R to truthfully reveal information. We can support R's truthful revelation if R's incentives for accepting over rejecting innovation are identical to D's for any m. This is true if $c_R = c_D/x$. I claim that D wants to induce R to conduct secret research. Here the relevant counter-factual is that D selects some other researcher, leading to a payoff of 0. D prefers to induce secret research if $\lambda[e_1 + \theta - c_D] - k_D x > 0$, which solves for condition 5. Finally, a cost profile must exist so that R wants to select secret research rather than deviate to either open research or scrapping the project (both yield expected value of 0). True iff, $\lambda[e_1 + \theta - x c_R] - k_R > 0$, as desired.

A.8 External ambiguity, and calibrating cost passing

We start with the baseline model. We then adjust it at only one decision node. If R asks for open research, that ask is continuous and thus there are many forms of open research. Specifically, at the first decision node (R selects between scraps, secret or open research), if R does not scrap the project, R's choice to research is represented by a continuous variable z. We allow R to set $z \in [0, 1 - x]$. If R sets $z = 0$, the model goes down the secret innovation pathway exactly as in the baseline model. If R sets $z > 0$, the model goes down the open research pathway with choices exactly as in the baseline model. However, we assume that the cost share parameter in the payoffs is adjusted, so that D accrues a $x + z$ share of the research cost if D approves research, and R accrues a $1 - z$ share of the research cost if D approves research.

Specifically, we only see a payoff adjustment under two conditions. First, if R asks for open research, D approves open research, and D rejects innovation, payoffs are: $U^D = -k_D(x+z), U^R = -k_R(1-z)$. Second, if R asks for open research, D approves open research, and D then approves innovation, payoffs are, $U^D = \pi + \theta - k_D(x+z) - c_D, U^R = \pi + \theta - k_R(1-z) - c_R x$.

Notice that $z = 1 - x$ is equivalent to the baseline payoffs from open research. But when $z < 1 - x$ R takes on a larger share of the burden from open research. If $z = 0$, the payoffs are the same as in secret research. Loosely, we can think about z as representing the expected chance that agents within the national security agency can keep the manager's knowledge of devilish details secret from some un-modeled, higher-level principal. This represents a case where D knows what R is doing, but R has informed D in such a way, that higher level principles may assign the blame to R. See the manuscript for more substantive motivation.

This variant of the model represents a tough theoretical test for the relevance of internal secrecy because only $x = 0$ represents true internal secrecy. We assume that the researcher is going to the manager for all $x > 0$, the manager fully understands what the research involves and can shut down a project if he wants to. We'll show that even under this tough test, conditions arise when the researcher still exploits internal secrecy.[33]

[33] We get even stronger results in favor of internal secrecy in a model where increasing z both increases the manager's cost, and probabilistic informs the manager of the devilish details. This would represent a setting where

First, we solve for the case where we observe cost sharing and partially external secret but internally open research.

Define $z^* = \min[1 - x, \frac{\lambda(e_1+\theta-c_D)}{k_D} - x]$.

Proposition 3 *If condition 2, 1 holds, and*

$$\frac{\lambda(e_1 + \theta - c_D)}{k_D} - x > z > 1 - \frac{\lambda(e_1 + \theta - c_R)}{k_R}$$

can be jointly solved for some $z \in (0, 1 - x]$, then the following strategies are SPE. R sets $z = z^*$ in the research phase. D's strategy is to accept research if $z \leq z^*$ and deny research otherwise. Regardless of how research occurs, D approves innovation if $\mathbb{E}[\pi|m] \geq c_D - \theta$, and reject innovation otherwise. Off path D rejects innovation absent research.

The extension does not adjust payoffs for innovation. Thus, as shown conditions 2, 1, guarantee that D will strictly reject innovation absent research, and reject innovation post-research if $\mathbb{E}[\pi|m] < c_D - \theta$ and accept it otherwise.

Turning to the choice to research. D does not make a choice if R selects secret research. If R selects a variant of open research, D approves research if: $\lambda(e_1 + \theta - c_D) - k_D(x + z) > 0$. This solves for the LHS of the equilibrium condition. Thus, when this condition is satisfied, D cannot profitably deviate to rejecting open research.

If R sets z too high, or scraps the idea, R's expected payoff is 0. R prefers to set z at some level D will accept over a choice that induces a payoff of 0 iff $\lambda(e_1 + \theta - c_R) - k_R(1 - z) > 0$. This solves for the RHS of the equilibrium.

We claim that if the equilibrium condition is satisfied, R sets z^*. z^* defines the the largest amount of cost-sharing that is both feasible (By assumption, bounded at $z \leq 1 - x$) and that D will accept. We've shown that R cannot profitably deviate to a higher z^* under stated conditions because D will reject. Since R' utility is increasing in D's responsibility, R cannot profitably deviate to sets $z < z^*$. This completes the proof.

As expected, R's incentives are to defray the political costs of research by passing them onto R's manager. This creates incentives for R to pursue open research over secret research $z = z^* > 0$. One might wonder, do researchers ever sustain internal secrecy from their managers if they have the option to pass on costs? We now identify the conditions where we still observe secret research,

Proposition 4 *If conditions 1, 2, 3, and*

$$\frac{\lambda(e_1 + \theta - c_D)}{k_D} < x \tag{10}$$

hold, then the following strategies are sub-game perfect. R selects secret research: $z = 0$. D's strategy is to deny all requests for research, deny innovation absent research, and approve innovation post-research if $\mathbb{E}[\pi|m] \geq c_D - \theta$ but not otherwise.

the research writes a vague report, or a very technical report where the devilish details are buried. In a situation like this, the manager may pick up the details but may not.

D will reject every open research request $z \in (0, 1-x]$, if $\lambda(e_1 + \theta - c_D) - k_D x < 0$. This gives us equilibrium condition 10. Thus, if R selects open research R's expected utility is 0. Note we are now in an identical situation to the baseline model. The result from proposition 1 carries through. We can support secret innovation if conditions 1, 2, 3 hold, as desired. Note condition 10 is easier to satisfy when c_D, k_D are high. This substantiates our claim in the manuscript that researchers only exploit informal briefs when the manager's costs are low, and that we expect to see this kind of informal briefing in the deep uncertainty pathway. However, we still expect true internal secrecy over the devilish details when condition 10 is satisfied.

A.9 Institutional design

We now introduce a higher-level principal who: (a) has a stake in the national security welfare of the country; (b) has the power to write the rules that govern how members of the executive incur costs. In the U.S. context, this principal could represent Congress.

We start with the monitoring extension presented in section A.6. In terms of timing, we add but one choice to the beginning of the game. We allow Congress to set $x \in [0, 1]$. All agents publicly observe x. At that moment, Congress becomes passive, and the game unfolds between R and D given the set x as it is presented in section A.6. Note, this framework closely matches how Congress writes rules for the national security community. Specifically, Congress pass general laws that determine the conditions under which a specific agent will face costs. These include laws that determine what actions are illegal, or constitute professional misconduct. It also includes who has a responsibility for their subordinates, and who has a responsibility to speak up if their managers abuse the law. Members of the intelligence community are then confronted with specific scenarios (e.g. the decision to pursue a particular idea) knowing what the laws that govern their actions are, the risks of exposure, etc.

As we shall see, setting x has two affects. First, it alters the strategic incentives of the agents in the research institution. Second, it imposes a direct cost on Congress because, consistent with our motivation that internal secrecy is important to sustain external secrecy, it raises the risk that foreign rivals will discover the programs and capabilities of our national security institutions.

We assume that Congress' utility function is similar to the manager's in that Congress incurs the research and innovation costs when the manager does. We assign c_O (O for overlord) as Congress's cost for pursuing innovation. We assume Congress suffers the common k, which is randomly drawn in this model and discussed in section A.6.[34] We allow the possibility that Congress suffers one additional cost, $g(x)$, which is weakly increasing in x and $g(0) = 0$. This cost represents the inevitable trade off between internal secrecy and external secrecy. As discussed in the concepts section, internal secrecy is what partly excuses agents from punishment when their team makes choices that they did not know about, or had limited ability to question. In an open institution $x = 1$, meaning that all agents are responsible for finding out what is happening in their own team and reporting wrongdoing when they see it. But as discussed in the concepts section, the higher x is, the greater risk there is that foreign rival will discover our intelligence practices. Putting these pieces together, Congress' utilities are:

[34] Note that since Congress takes the first action, Nature has not yet drawn k when Congress acts. It does not matter of Congress observes k or not, because Congress has no additional actions.

$$U^O(\text{research, innovation}) = \pi + \theta - k - c_o - g(x)$$

$$U^O(\text{no research, innovation}) = \pi - c_o - g(x)$$

$$U^O(\text{research, no innovation}) = -k - g(x)$$

$$U^O(\text{no research, no innovation}) = -g(x)$$

The theoretical concern that motivates this extension is as follows. Even if it is true that a high amount of internal secrecy would incentivize agents to participate in the don't-ask-don't-tell scenario, Congress would anticipate this concern and change the institutional rules so that national security agents would not exploit it. Thus, our goal is to show that conditions exist where Congress would prefer to live with don't-ask-don't-tell, rather than prevent it.

We proceed as follows. First, we focus our analysis on conditions where we can induce don't-ask-don't tell for $x < x^*$, but Congress can prevent this behavior by setting $x \geq x^*$. Second, we isolate the two conditions—$x < x^*$, $x \geq x^*$—and separately solve for strategy profiles for R and D that we can support on path in each case. Along the way we identify Congress' utility from setting x in either range, given R and D play these strategies. In the $x \geq x^*$ case, we show that Congress induces D to monitor if ever D does not observe research. This creates the following test for our analysis. If O ever sets $x < x^*$ in equilibrium, then we can say that O is not willing to set x sufficiently large to prevent agency loss. Third, we characterize an equilibrium. Finally, we solve for a minimum condition where Congress has a profitable deviation from every $x \geq x^*$ to $x = 0$, given the strategies for R and D that x will produce. Thus, we identify conditions where Congress would not set x large enough to scuttle don't-ask-don't-tell because Congress has at least one profitable deviation to $x = 0$.

A.9.1 Parameter restrictions

To start, we focus on fixed set of values that allow us to support the don't-ask-don't-tell equilibrium defined in proposition 2 for a range of x. Using the same definition for $\mathbb{E}[k|sr, nor]$ as above, define $x^* = \frac{\lambda[e_1+\theta-c_D]}{\mathbb{E}[k|sr,nor]}$. This is the value of x for which condition 4 becomes an equality.

Then, define a set of fixed values of, $p(), \sigma, f(k), c_R < c_D, \theta$ for which we can support the don't-ask-don't-tell equilibrium defined in proposition 2 for all $x \in [0, x^*]$. Note that x appears in conditions 3 4, and both are easier to satisfy as x decreases. In the limit, at $x = 0$, condition 8 is always satisfied, and condition 3 reduces to $k < \lambda(e_1 + \theta)$. Finally, $\bar{k} \to \lambda(e_1 + \theta)$.[35] Also note that conditions 1 and 2 do not depend on x, and are assumed satisfied by our parameter restriction.

Summing up the implications of these restrictions for R and D's strategy. By construction, if $x \geq x^*$ we cannot support proposition 2 because condition 4 is violated. But if $x < x^*$ condition 4

[35] While it is true that adjusting x influences $\bar{k}$ which increases $\mathbb{E}[k|sr]$. It is also the case that $\mathbb{E}[k|sr]$ is a real valued number for any x, thus, if $x = 0$, inequality 4 is always satisfied.

holds, and we can support no monitoring and secret innovation. As a result, Congress can induce don't-ask-don't tell if Congress sets $x < x^*$, but will prevent it otherwise.

In what follows we separately analyze the $x < x^*, x \geq x^*$ cases. Where not specified, we define strategies for R and D we can support in a PBE given x that falls into these respective ranges. We also specify O's expected utility given those strategies.

A.9.2 The $x < x^*$ case

By construction, we can support the strategy for R and D as stated in proposition 2.

Remark Conjecture that if Congress sets $x < x^*$ that R and D play the strategies described in proposition 2. Then, Congress's expected utility for setting $x < x^*$ is:

$$EU^O(x < x^*) = pr[k < \lambda(e_1 + \theta - c_R x)]](\lambda(e_1 + \theta - c_O) - \mathbb{E}[k|k < \lambda(e_1 + \theta - c_R x)]) - g(x)$$

Note we cannot make strong claims about which $x \in [0, x^*)$ maximizes Congress' utility because Congress faces a three-way trade-off between increasing the direct cost, $g(x)$, increasing the probability that R will pursue research, which increases the value that Congress expects to accrue because more profitable programs get funded, and increasing the expected cost that Congress incurs should research happen. It is possible that there are multiple maximum values, and they may take on the corner $x = 0$.

A.9.3 The $x \geq x^*$ case

We now characterize one strategy profile for R and D that we can support on path, under the assumption that $x \geq x^*$.

Lemma 6 *Fixing Congress's strategy at $x \geq x^*$, then we can support the following strategies on path in a PBE.*

- *D always monitors if D fails to observe open research. Regardless of how D comes to identify research, D approves research if $k \leq \lambda[e_1 + \theta - c_D]$ and rejects it otherwise. D rejects all innovation absent research, and approves innovation post-research iff $\mathbb{E}[\pi|m] > c_D - \theta$.*
- *R requests open research if $k \leq \lambda[e_1 + \theta - c_D]$ and scraps the project otherwise.*

Because condition 1 and 2 are satisfied and constant for any x, we know that D will reject innovation absent research, D will approve innovation iff $\mathbb{E}[\pi|m] > \theta - c_D$, and D will reject open research if $k > \lambda[e_1 + \theta - c_D]$. What is more, we can still use the same definitions for e_0, e_1, λ, which do not depend on x.

We conjecture that if $k > \lambda[e_1 + \theta - c_D]$, R always scraps the project, and D always monitors. Further, if after monitoring, D did observed secret research, D would reject research. At the moment R decides to scrap the project, R's expected utility from on path play is 0. At the moment, D decided to monitor, D's expect utility from on-path play is 0.

Starting with D's incentive to reject secret research if discovered. As shown in proposition 2, D's reject research post-monitoring if $k > \lambda[e_1 + \theta - c_D]$. Thus, D cannot deviate from rejecting research if D monitors and discovers that R has conducted secret research. Turning to D's incentive to monitor. Note R does not play secret research on path. Thus, if D does not observe research, D expects 0 from not monitoring. Thus, D is indifferent between monitoring (on path) or not. Turning to R's incentive to scrap the project. As shown in proposition 2, if R deviates to open research, D rejects open research (and then innovation) if $k > \lambda[e_1 + \theta - c_D]$. This leaves R indifferent between scrapping the project and selecting open research. If R selects secret research, R expects $-k$ given that D always monitors and shots down research. Clearly, R does worse from deviating to secret research.

We conjecture that if $k < \lambda[e_1 + \theta - c_D]$, R always requests open research and D approves. If R deviated to secret research D would monitor and approve. Thus, R's expected utility from on path play, at the moment R requests open research is $\lambda[e_1 + \theta - c_R x] - kx$, D's expected utility at the moment D approves open research is: $\lambda[e_1 + \theta - c_D] - k$.

As shown in proposition 2, D prefers to accept open research to not if $k < \lambda[e_1 + \theta - c_D]$. If R deviates to secret research, D observes no research. As just shown, D always monitors given this observation. But in this off-path case, D's monitoring discovers secret research and $k < \lambda[e_1 + \theta - c_D]$. As just shown, D would approve. Thus, R is indifferent between secret and open research. Finally, consider R's deviation to no research. In this case, R gets 0. R can only profitably deviate to scrapping the idea if, $k < \frac{\lambda[e_1 + \theta - c_R x]}{x}$. This is always satisfied if $c_R < c_D$ (true by the construction of the scenario) and also $k < \lambda[e_1 + \theta - c_D]$. To see it, set $x = 1$, and R cannot profitably deviate if, $k < \lambda[e_1 + \theta - c_R]$. Thus, R's incentive to deviate does not impose an additional parameter restriction.

Summing up, we've solved for a strategy profile for R and D that we can support as part of a PBE given $x \geq x^*$. While this is not the only strategy profile we can support, it is the one that can guarantee no agency loss at the lowest level of x. Thus, it is important to focus on it because (a) it allows Congress to avoid agency lost at the lowest $g(x)$, and (b) x only enters into Congress' payoff through $g(x)$. Thus, Congress strictly prefers $x = x^*$ over $x > x^*$.

Remark Suppose that in an equilibrium if Congress sets $x \geq x^*$, that Congress induces R and D to play the strategies described in Lemma 6. Then Congress's expected utility at the moment Congress sets $x \geq x^*$ is:

$$EU^O(x \geq x^*) = pr[k < \lambda(e_1 + \theta - c_D)](\lambda(e_1 + \theta - c_O) - \mathbb{E}[k | k < \lambda(e_1 + \theta - c_D)]) - g(x) \quad (11)$$

A.9.4 Equilibrium

Define $\tilde{x}$ as the largest[36] x that maximizes:

[36] It may not be unique, but that is not the point. We pick the largest x because our goal is to show that D will pick one that is less than x^*.

$$\begin{cases} EU^O(x \geq x^*) & if\ x \geq x^* \\ EU^O(x < x^*) & if\ x < x^* \end{cases}$$

We now conjecture a set of strategies. O plays $\tilde{x}$. If $x \leq x^*$ R and D play the strategies written in proposition 2. If $x > x^*$, then R and D play the strategies described in Lemma 6.

Proposition 5 *Under our parameter restrictions, the conjectured strategies form a PBE.*

In section A.9.3 we showed that we could support R and D's strategy given an observed $x \geq x^*$. In section A.9.2 we argued via reference to proposition 2 that we could support R and D's strategy given an observed $x < x^*$. In the respective sections we defined O's expected utility from setting x given that it would induce the respective strategy. What is not proven is that O cannot profitably deviate from playing $\tilde{x}$, given the strategies for R and D it will induce. But by construction, $\tilde{x}$ is the x that (weakly) maximizes O's utility. Trivially O cannot deviate from it.

A.9.5 When will O set x to induce don't-ask-don't tell

To be clear, this result does not specify what $\tilde{x}$ is. It is possible that O would always set $\tilde{x} \geq x^*$. Our central claim is that conditions exist where we cannot support $\tilde{x} \geq x^*$. Thus, our final task is to verify that conditions exist where O will set $x < x^*$ and thus induce R and D to play the don't ask don't tell behavior.

We do so in two steps. First, we establish the best O can do if O sets x so large as to prevent don't-ask-don't-tell. Second, we establish conditions where O has at least 1 profitable deviation from O's best $x \geq x^*$.

Remark $EU^O(x \geq x^*)$ is maximized at $x = x^*$ for $x \geq x^*$. This yields:

$$EU^O(x = x^*) = pr[k < \lambda(e_1 + \theta - c_D)](\lambda(e_1 + \theta - c_O) - \mathbb{E}[k|k < \lambda(e_1 + \theta - c_D)]) - g(x^*) \quad (12)$$

Note that Congress's total expected utility for setting $x \geq x^*$ is weakly decreasing in x, because Congress must pay $g(x)$, But R and D's strategy are invariant to x, as are other features of Congress' utility. It follows that if Congress sets x to prevent don't ask don't tell, Congress sets $x = x^*$.

We now show conditions exist where Congress can profitably deviate from $x = x^* \rightarrow x = 0$. To be clear, this does not mean that $\tilde{x} = 0$. But it does guarantee that $\tilde{x} < x^*$, which is the point of our analysis. We focus on $x = 0$ because it simplifies the boundaries $\bar{k}$, allowing for a clear comparison. In particular, Congress's expected utility from $x = 0$ is:

$$EU^O(x = 0) = pr[k < \lambda(e_1 + \theta)]](\lambda(e_1 + \theta - c_O) - \mathbb{E}[k|k < \lambda(e_1 + \theta)]$$

For emphasis, we re-write it as:

$$EU^O(x=0) = EU^O(x=x^*) + g(x^*) \\ + pr[k \in [\underline{k}, \lambda(e_1+\theta)]] \times (\lambda(e_1+\theta-c_O) - \mathbb{E}[k|k \in [\underline{k}, \lambda(e_1+\theta)]]) \quad (13)$$

Remark In equilibrium, we cannot support $\tilde{x} \geq x^*$, on path if $EU^O(x=0) > EU^O(x=x^*)$:

$$g(x^*) > pr[k \in [\underline{k}, \lambda(e_1+\theta)]] \times (\mathbb{E}[k|k \in [\underline{k}, \lambda(e_1+\theta)]] - \lambda(e_1+\theta-c_O))$$

We note two facts about this inequality. First, if $g(x^*)$ is large, Congress will strictly prefer complete internal secrecy that induces don't ask don't tell to setting $x = x^*$. Thus, it instantly follows that the concern over external secrecy alone can drive Congress to set $x < x^*$.

But also notice that if we can ignore the direct costs by setting $g(x^*) = 0$ the inequality can still hold if:

$$\lambda(e_1+\theta-c_O) > \mathbb{E}[k|k \in [\underline{k}, \lambda(e_1+\theta)]]$$

The LHS of this inequality captures that lowering x from x^* to 0 means that research will happen leading to more innovation, and this raises the chance of welfare enhancing innovations. The RHS of this inequality captures that lowering x means that the additional research comes at a higher level of political costs.

B Monitoring the Soviets and the origins of U2

The main paper examined two cases of innovation: the search for mind control and the origins of the reconaissance satellite. This section examines a third case, the origins of the U-2 spy plane. As will be described in detail, this case provides additional inferential leverage that further validates the theory.

One of the United States' most pressing priorities in the early years of the Cold War was gaining better understanding of the Soviet Union's capabilities.[37] Without it, there was a heightened risk of insecurity, the possibility of arms racing, and even inadvertent war. But an aggressive and capable air defense made the prospect of overflights below a certain altitude a risky endeavor. Thus, the search for a high-flying reconnaissance aircraft was on.

The initial effort was spearheaded by the Air Force and various affiliated organizations. One of the most notable efforts was spearheaded by the Wright Air Development Command led by Major John Seaberg. In March 1953, Seaberg settled on desired specifications for the aircraft. He wanted it to "have an optimum subsonic cruise speed at altitudes of 70,000 feet or higher over the target, carry a payload of 100 to 700 pounds of reconnaissance equipment, and have a crew of one" (Pedlow and Welzenbach, 1992, 8). Seaberg solicited proposals from a number of smaller airframe manufacturing companies. He was seemingly interested in any solution that met his specifications and believed

[37] https://nsarchive2.gwu.edu/NSAEBB/NSAEBB74/U2-02.pdf.

smaller companies would take the project more seriously and move more quickly (Pedlow and Welzenbach, 1992, 8). He heard four bids:

- Fairchild Engine and Airplane Corporation proposed a single-engine aircraft, the M-195, which promised to reach a maximum altitude of 67,200 feet.
- Bell Aircraft Corporation proposed a twin-engine plane, the Model 67, or later the X-16, which promised to reach 69,500 feet.
- Glenn L. Martin Company proposed "a big-wing version of the B-57 called the Model 294, which was expected to cruise at 64,000 feet."
- Lockheed Aircraft Corporation proposed a modified, single engine aircraft that approximated sailplane, the CL-282, which promised to reach just north of 70,000 feet (Pedlow and Welzenbach, 1992, 9).

In a moment we will support our theory by examining who funds what and why. Before that, we emphasize the unique features of this case that help us validate our core counterfactual claim.

B.1 Counter-factual reasoning at this unique period in history

Our theory is built on a counter-factual claim: secret institutions pursue research that more open institutions would reject. This is difficult to validate in the modern institutional context for three reasons. First, the military and intelligence organizations employ many scientists who devise ideas on their own. When a CIA scientist conceives of a novel idea and explores it, for example, we cannot know whether the military would have rejected it. Second, scientists and engineers select into the institutions they work for. As such, we cannot know if CIA scientists are similar to military scientists and vice versa. Finally, private companies that devise new ideas know they can pitch them to highly secret parts of the government like the CIA through classified contract mechanisms. If our theory is right, we may never observe them take ideas to the military.

A confluence of factors in this case provides a unique opportunity to test our theory. First, the companies that bid on reconnaissance aircraft all believed that the Air force was effectively the sole outlet for such pitches.[38] Interestingly, however, a relevant secret organization did exist. In July 1954, President Eisenhower tapped the President of MIT, James Killian, to head a group of scientific experts called the Technology Capabilities Panel (TCP) (Richelson, 2002, 11). Its existence was not widely known: "As with other secret panels formed by chief executives to deal with intelligence matters, Congressional input was missing from the TCP deliberations and few Congressmen knew it even existed, although many of its decisions had an immense impact on the nation's military and intelligence preparedness" (Laurie, 2001, 5).

Project Three, one of three entities comprising the TCP, was a small group broadly focused on intelligence capabilities. It was not specifically tasked with developing proposals for overhead reconnaissance aircraft. Thus, the small and secretive Project Three members were not soliciting bids for such aircraft, and nobody expected that they would. However, the extreme secrecy that

[38] Although the CIA had developed several branches to deal with scientific intelligence and research and development in the early- to mid-1950s, they did not have much experience at that time with technical collection systems. See Fischer (2001).

surrounded Project Three meant that they could develop research ideas in small teams that outsiders would not know about. Thus, unlike the Air Force, they exhibit the internal secrecy that our theory requires for secret innovation.

Based on this context, it is reasonable to assume that the Wright Air Development Command and any other relevant Air Force-related entity would hear all bids pertaining to overhead reconnaissance and had first right of refusal. Moreover, any project they did fund would at least be scrutinized by the broader Air Force leadership and possibly Congress. They would have also likely believed that anything they rejected would not be funded. However, as just noted, Project Three was quietly lurking in the background and ready to pick up rejected proposals if they so chose. This allows us to evaluate our counterfactual because we can observe: (1) what the open institution actually chose to accept and reject and; (2) given what the open institution rejected, what the secret institution chose to accept and reject.

B.2 Who funds what and why

The Air Force opted to pursue two proposals, the modified version of the B-57 from Martin which was viewed as a short-term solution and the Bell X-16 which promised better results in the medium-term. Bell was contracted to produce 28 such aircraft. At the same time, the Air Force rejected the Fairchild and Lockheed proposals. The Fairchild proposal was relegated to the dustbin of history. The Lockheed proposal was not. Lockheed took their proposal to various parts of the Air Force—including the Wright Air Development Command as well as Strategic Air Command and the Office of Development Planning—all of whom rejected it (Pedlow and Welzenbach, 1992, 11-12). Along the way, Project Three members learned of the Lockheed proposal and were immediately interested in it (Pedlow and Welzenbach, 1992, 31). As we will detail more in a moment, they undertook intense secretive research into CL-282's viability and verified that it would work. This project was later handed to the CIA as the U-2 project.

We predict that open organizations facilitate innovation when the benefits are clear (e_0 is positive), and there is not much disagreement about the likely effects (σ_0 is low); they will reject ideas that are radically new because they know little about them. Even though new ideas could have benefits, they could also cause damage. Open institutions are unlikely to take on projects like this even in the research phase (e_0 is near 0). Of these ideas, we predict that secretive institutions will pick them up as research projects if the potential outcomes vary widely (σ_0 is high). That is, there is a risk of catastrophic damage towards mission objects and enormous benefits that extend beyond what the other proposals could accomplish.[39]

This is precisely what we find. The Air force funded two safer projects that incrementally advanced the state of overflight. The modified B-57 is an obvious example. The goal was to "improv[e] the already exceptional high-altitude performance of the B-57 Canberra" (Pedlow and Welzenbach, 1992, 9). It "featured lengthened wings, accommodations for cameras and sensors, and uprated twin engines" (Merlin, 2015, 1). The Bell X-16 was slightly more advanced than the B-57. The modifications made to reduce weight and reach higher altitudes were far less radical than the CL-282 (Merlin, 2015, 4-5).

The U-2 was radical by design. Senior Lockheed designers prioritized "nonstandard" elements,

[39] That is, there is uncertainty about whether the innovation will move the U.S. closer to or further from its policy objectives.

including "the elimination of landing gear, the disregard for military specifications, and the use of very low load factors" (Pedlow and Welzenbach, 1992, 10). Several elements of what was eventually dubbed the CL-282, and would later become the U-2, "were adapted from gliders. Thus, the wings and tail were detachable. Instead of conventional landing gear," Kelly Johnson, the lead developer, "proposed using two skis and a reinforced belly rib for landing—a common sailplane technique—and a jettisonable wheeled dolly for takeoff." As a declassified history of the U-2 puts it, "Essentially, Kelly Johnson had designed a jet-propelled glider" (Pedlow and Welzenbach, 1992, 12).

Part of Seaberg and the Wright Air Development Command's rationale for rejecting the CL-282 proposal speaks to their uncertainty about whether it would work. Seaberg pointed to its use "of the unproven General Electric J73 engine. The engineers at Wright Field considered the Pratt and Whitney J57 to be the most powerful engine available." All three of the other proposals they received from small manufacturers relied on the latter. Moreover, Seaberg and colleagues viewed "[t]he absence of conventional landing gear" on the CL-282 as a "shortcoming." Because the other proposals, including the most promising—the Bell—had "normal landing gear," they were considered "more conventional aircraft" (Pedlow and Welzenbach, 1992, 12-15).

Other Air Force commands also registered dismay at the novel features of CL-282. General Curtis LeMay, the head of Strategic Air Command, apparently "stood up halfway through the briefing, took his cigar out of his mouth, and told briefers, that if he wanted high-altitude photographs, he would put cameras in his B-36 bombers and added that he was not interested in a plane that had no wheels or guns." He called the meeting "a waste of his time" (Pedlow and Welzenbach, 1992, 12).[40]

According to the declassified history of the U-2, another driving factor in the Air Force's rejection of the CL-282 had to do with their "preference for multi-engine aircraft." This was based on familiarity and their experience with multi-engine aircraft during World War II and likely explains why they also opted for the Bell and Martin designed but rejected the Fairchild bid, which relied on a single engine. Moreover, "aerial photography experts" at the time "emphasized focal length as the primary factor in reconnaissance photography and, therefore, preferred large aircraft capable of accommodating long focal-length cameras" (Pedlow and Welzenbach, 1992, 13)

As the foregoing makes clear, the CL-282's novel design meant that many in the Air Force were skeptical about its chances of success. In terms of the model's parameters, the balance of Air Force staff thought the overall impact of the project would cause no benefit (or harm) for surveilling the Soviet Union and ultimately ensuring peace. However, some raised concerns which implied that it could have catastrophic effects: "there was the feeling shared by many Air Force officers that two engines are always better than one because, if one fails, there is a spare to get the aircraft back to base... Furthermore, a high-altitude reconnaissance aircraft deep in enemy territory would have little chance of returning if one of the engines failed, forcing the aircraft to descend" (Pedlow and Welzenbach, 1992, 13). In other words, there was concern that a single-engine plane that was missing key parts could crash inside the Soviet Union and conceivably spark a conflict.

To be sure, not everyone in the Air Force shared the view that the Bell and Martin proposals were superior to the CL-282. Trevor Gardner, Special Assistant for Research and Development, and some other officials thought it had potential. They believed "it gave promise of flying higher than the other designs and because at maximum altitude its smaller radar cross-section might make it

[40] LeMay's reaction illustrates one way that military culture imposes costs on innovators. As we argued, this makes innovation difficult in open institutions.

invisible to existing Soviet radars" (Pedlow and Welzenbach, 1992, 15). Thus, if it worked, its value would be larger than the other projects.

Taken together, these divergent views support the notion that there was deep uncertainty about what CL-282 would accomplish. While some believed it was unlikely to work and therefore have no effect, others thought it could have either very negative or very positive (i.e. more positive than the other designs) effects. If the Air Force had been the only organization that could have considered the overflight proposals, one of the most important innovations of the twentieth century may never have seen the light of day (Pocock, 2000, 14).

Project Three members were themselves sensitive to the risks associated overflight over the Soviet Union.[41] But despite these risks, they pursued the project because of the enormous potential upside if the project was successful. "By the end of October [1954], the Project Three meetings had covered every aspect of the Lockheed design. The CL-282 was to be more than an airplane with a camera, it was to be an integrated intelligence-collection system that the Project Three members were confident could find and photograph the Soviet Union's Bison bomber fleet and, thus, resolve the growing 'bomber gap' controversy." They were also taken with the prospect that the proposal could be "the platform for a whole new generation of aerial cameras" (Pedlow and Welzenbach, 1992, 31).

Their approach to research supports our theory in two additional ways. First, they operated in secret. Land and his team "began developing it into a complete reconnaissance system," meeting in small-group settings with usually less than 10 people present. Second, they did not instantly recommend production of U-2 planes. Rather, they exploited secrecy to determine if the project was viable. Once they realized it was, they revealed what they had been doing to the CIA Director and to President Eisenhower who was extremely receptive. He "approv[ed] the development of the system, but ... stipulat[ed] that it should be handled in an unconventional way so that it would not become entangled in the bureaucracy of the Defense Department or troubled by rivalries among the services" (Pedlow and Welzenbach, 1992, 33).[42]

Interestingly, the project also helped the TCP realize that secret organizations like the CIA were well-suited to the task of overseeing radical innovations of this kind. As the TCP argued to CIA Director Allen Dulles in a memo, "this seems to us the kind of action and technique that is right for the contemporary version of the CIA; a modern and scientific way for an Agency that is always supposed to be looking, to do its looking. Quite strongly, we feel that you must always assert your first right to pioneer in scientific techniques for collecting intelligence... This present opportunity for aerial photography seems to us a fine place to start" (Land, 1954*b*).

C National Security and Innovation Literature

Since our theoretical framework is closest to principal-agent theories of organizational innovation, we focus our review on that literature. We also review works in international relations and bureaucratic politics that help us justify changes in our assumptions. However, our paper has broad substantive interest for scholars of innovation and security broadly defined. Here we review four different

[41] See Land (1954*a*).

[42] Interestingly, the Air Force eventually comes around to accepting the proposal but does not actually abandon their X-16 program until the U-2 was operational.

strands of this literature, explain how we connect and contribute to them:

1. Bureaucracy and barriers to and opportunities for military innovation;
2. Adaptation and military innovation;
3. Conflict processes and innovation, which can examine autocratic repression or terrorism and innovation;
4. The strategic implications of new technology.

Many of the concepts we describe intersect with these literatures. But we frequently arrive at surprising conclusions for all of them. In what follows, we explain how our theory intersects with these important literatures and clarify differences.

C.1 Barriers and opportunities for military innovation

A large literature in security and strategic studies examines military innovation. Many of these analyses begin with the premise that, despite the importance of innovation to national security, military innovation is rarer than we might expect it to be. Why? The answer, in brief, is that innovators face costs of different kinds. One common impediment is that militaries are "hierarchical, inflexible, and rigid" (Jungdahl and Macdonald, 2015, 467). As Grissom (2006, 919) argues in his review of this literature, most scholars argue that "military organizations are intrinsically inflexible, prone to stagnation, and fearful of change." What this means in practice is that individuals are often reluctant to suggest new ideas for professional or cultural reasons, and new ideas that do get proposed can often get shut down.

Despite these barriers, militaries sometimes innovate. Thus, another key task of this literature is to answer the following question: what explains how militaries can overcome bureaucratic inertia or military culture to innovate? Some argue that military organizations may innovate when they face external pressures from the outside, usually from civilians (Posen, 1984). Another is when senior members of the military re-conceptualize their tasks and create career paths for new officers that incentivize the embrace of this new way of thinking (Rosen, 1988). A third set of explanations focuses on cultural differences (Adamsky, 2010; Farrell and Terriff, 2002) According to one study, a "receptive culture" can facilitate new thinking and vice versa.[43] A fourth argues that innovation requires special incubators where individuals can collaborate, try out ideas, and push the envelope Jensen (2016). There are others (Grissom, 2006).

While each of these pathways are distinct in important ways, they all share a common strategic logic. First, individuals inside the military face barriers (i.e. costs) to innovation. Therefore, they either do not voice their ideas, or are unable to push their ideas through the military bureaucracy. This explains why innovation does not happen often. Second, opportunities for innovation arise when military leaders, or outsiders with power create incentives (i.e. lower the costs associated with pursuing innovation). Things like new pathways to promotion, visionary civilians that intervene to support and defend new ways of doing business, and incubators where individuals can test

[43] Price (2014). Lee (2019) has shown, for example, that the Air Force's cultural preference for manned systems led it to reject innovations in drone technology for longer than would otherwise be the case if one were using a strictly rationale model.

ideas outside the formal process are a way for would-be innovators to safely conceive of ideas, develop them, and potentially implement them without incurring significant costs. Without these cost-lowering mechanisms, the argument goes, innovation does not happen.

Our theory accounts for these conditions in the costs and benefits parameters. The logic of our model under a specific set of parameters is consistent with the logic of these arguments. We find that researchers will not *openly* pursue innovation even when the policy implications are important (the expectation of pi is positive) if the organization imposes large personal costs on the agents.

The critical difference between our theory and this literature is what happens when the costs and benefits are high. Scholars of military innovation typically argue that if the costs of pursuing research are high then the innovators simply do not pursue their ideas. As noted, their logics for military innovation largely follow a similar process: some kind of organizational change transpires that lowers the costs associated with agents openly pursuing innovation; the researcher realizes that the organization is accommodating of new ideas; the researcher then raises their ideas with their manager so that they can openly pursue them. In our theory, national security researchers sometimes face another option: secret innovation. Rather than taking their idea to their manager, or sharing it broadly with others in their organization, a small team of researchers can pursue an idea in secret. This gives the researcher autonomy to pursue their idea and demonstrate its plausibility. It also allows different agents to distribute the high institutional costs associated with pursuing new ideas.

In this way, our theory illuminates that existing studies emphasize open, national security innovation in the way that we define openness.[44] As written in the manuscript, open research refers to a setting where individuals broadly share their ideas with their managers, people with budgetary oversight, and many others across their organization and possibly outside their organization.[45] What is more the costs that these scholars describe usually stem from openness. Consider that bureaucratic inertia, or cultural barriers only prevent pilot testing if ideas are shared openly. If a small team of researchers does not ask permission, they do not face bureaucratic inertia.

There are several other ways in which our theory differs from, but complements, broader literature on military innovation which includes both doctrinal innovations as well as technological and tactical innovations (Beard, 1976; Jungdahl and Macdonald, 2015; Sapolsky, 1972) First, most of these accounts emphasize innovation that occurs through a top-down process. Our focus entails a heavy bottom-up component (Griffin, 2017, 214).[46] Second, much of this scholarship on military innovation has a bias towards *successful* innovations.[47] By focusing on the process or pursuit of innovation, our study allows for the prospect that many of these ideas, particularly those pursued

[44] Scholars such as Kurth Cronin (2020, 23-28) discuss "closed innovation," defined as "state organizations creat[ing] and control[ing] high-end military technologies" such as nuclear weapons. Even in this case, though, while innovation may be hidden from the *outside* world it is still open internally within the government.

[45] Although they do not usually describe it this way, the existing security studies literature usually focuses on open innovation under this definition. Perhaps the clearest example of this is innovations in doctrine, a common focus of this literature. When doctrinal innovation happens, it is usually carried out in broad view of many parts of the military. It requires many services and branches to work together. Even during periods of conceptualization, new doctrine requires combat experts to interface with logistics, strategic intelligence, manpower and budget experts, defense contractors, and more. Moreover, since new doctrine requires new field manuals, soldiers tend to find out important details of doctrine as it is being developed.

[46] For exceptions, see Jungdahl and Macdonald (2015); Kollars (2014).

[47] This is evidenced by the way many scholars define innovation, which often requires things like improvements in military effectiveness. See Grissom (2006, 907). As Posen (1984, 29) notes, however, "Neither innovation nor stagnation ... should be valued a priori.

in secret organizations, will fail.

C.2 Adaptation and military innovation

A second literature examines diffusion and adaptation. This is similar because it examines military innovation. However, they focus on how existing military technologies diffuse cross-nationally. Horowitz (2010, 3), for example, develops the "adoption-capacity theory" to explain "why some military innovations spread and influence international politics while others do not, or do so in very different ways." In a somewhat similar vein, Gilli and Gilli (2019, 141) examine the logic of imitation, asking whether America's rivals can "easily imitate its most advanced weapon systems and thus erode its military-technological superiority."

The aspect of these studies that is most similar to ours examines different ways that states adopt the same technology. This could be thought of as tactical innovations. However, these tactical innovations are typically described as open, and the primary barriers is in adopting an existing technology and not in finding new ways to use it.

C.3 Innovation among autocrats and terrorist groups

Our framework also differs from a newer literature on innovation among terrorist organizations and autocratic regimes. Regarding terrorist groups, innovation is often driven by the need to evade a target's defenses, amplify lethality, and shape public opinion (Horowitz, Perkoski and Potter, 2018). The precise characteristics of terrorist organizations, their leaders, and their broader environment, however, shape whether they are successful.[48] One of the key differences between these studies and our own is that terrorist organizations as a whole are insensitive to the costs of innovation whereas the individuals in our model are political actors and researchers with an entirely different incentive structure.[49]

Finally, there is an emerging literature that examines innovation and autocratic regimes. A key focus of these works is how dictators can exploit technological innovations to their advantage. This includes the use of the Internet and other technologies for the purposes of repression and surveillance (Dragu and Lupu, 2021; Gohdes, 2020). In these studies, autocratic leaders are exploiting existing technologies that may have been developed with an entirely separate purpose in mind for their own ends, including regime survival and population control. Like terrorist organizations, they are also insensitive to costs. As noted, our focus is on the sources of innovation in a situation where there are political actors who can distribute costs to subordinates.

C.4 Strategic implications of emerging technology

A growing litreature emphasizes the strategic implications of emerging technology (see Sechser et al., 2019, for review). We partly use this litearture to justify our claim that the benefits of innovation (i.e. whether innovation moves you closer or further from your policy goals) is uncertain. This

[48] See Moghadam (2013); Perkoski (2019).

[49] To be sure, terrorists may be sensitive to how the public will *perceive* an innovation such as suicide bombing but are themselves by and large insensitive or at least willing to incur enormous costs given the nature of asymmetric conflict.

literature is more about what states do with innovations once they have them. It is less about why states decide to pursue them in the first place (Garfinkel and Dafoe, 2019; Horowitz, 2016; Zhang et al., 2021).

D Principal-agent literature

Our substantive focus is foreign policy and international relations. However, as we discuss in the manuscript, the structure of our theory is closest to principal-agent theories of organizational innovation in the private sector (Lai et al., 2009). These theories emphasize aspects of PA problems not commonly studied by international relations scholars. In what follows, we explain how our theory fits within the PA framework. We then clarify important differences with three applications of PA theory in IR.

D.1 What makes our theory a principal-agent theory?

PA theory is very broad (Eisenhardt, 1989). There are many types of principal-agent problems that scholars study including moral hazard, agency loss, adverse selection, credible communication, and unjust reprisals (Stiglitz, 1989; Hart and Holmström, 1987). While each problem is different, they are united by a few common elements. In this section, we describe the elements of a PA theory and how our theory includes these elements.

A basic principal-agent dynamic (or contract theory) involves at least one agent and at least one unified principal that have asymmetric preferences and in which the agent is given a choice to impact the principal's welfare (Miller, 2005). Our basic institution models these elements. We study a researcher and manager who vary in their cost functions. As a result of these cost functions, situations arise where the researcher wants to pursue research and development but the manager does not. We make one assumption that is common in models of innovation: the effects of pursuing a policy follow from imperfect information and are not known to either player. This assumption is not common in PA models of policymaking (e.g. Downs and Rocke, 1994). The reason is that policymakers (i.e. the agent) knows whether their choice will benefit the principal with a large degree of confidence (i.e the public); at least *ex post*.

Beyond this difference, we make a novel assumption in the basic model that departs from PA models of innovation: the researchers can exploit secrecy to distribute costs. This creates a dynamic in which the researcher can incur costs to pursue outcomes that the manager would veto. We study the impact of this additional assumption under complete information because it generates a novel tension not typically appreciated in PA models.

Principal-agent theories introduce problems through asymmetric information, and a principal's initiative (Miller, 2005). The specific type of principal-agent problem varies depending on how scholars introduce private information (Hart and Holmström, 1987). We model two variants of a principal-agent problem. The first appears in the monitoring extension that supports Proposition 2. The second is present as a trust in the researcher section that supports Lemma 2. The first represents a monitoring problem, the second represents a credible advice problem. Past scholars examine how variation in the costs of monitoring, agent selection, or punishments can elicit agency compliance and the credible revelation of information. However, we find that secrecy paradoxically

alleviates many of the common problems of asymmetric preferences and information. It also creates new incentives for the manager to extract value from the researcher's compliance.

D.2 How is this different from PA models in international relations and foreign policy studies?

Here we describe three literatures that examine principal-agent problems in international relations: hierarchy, security force assistance, and gambling for resurrection.

We start with a joint-discussion of hierarchy (Hawkins et al., 2006; Nielson and Tierney, 2003; a. Lake, 2001) and security force assistance (Biddle et al., 2018; Ladwig, 2016). Of course, these empirical domains are very different from each other. Further, each domain includes many different studies that tackle different aspects of the PA problem. However, they are all united by the fact that they assume the principal and agent come from different states and therefore have dramatically different preferences. Scholars of security force assistance assume that the principal is either US military advisers or the entire US military and the agent is the military of another state (e.g. the Afghan army).

We do not focus on a situation like this. Consistent with organizational models of principal-agent theory and innovation, we examine individual employees (or small groups of individuals) who work at a single organization (or a handful of closely connected agencies that share a common mission within the executive branch of a single country; like the CIA and NRO). To match this domain, we assume that the researcher and manager both share an interest in advancing the organization's overall goals (both researcher and manager's utility is increasing in π). However, their preferences over research and development still vary because the personal and professional incentives of managers and researchers vary (c, k can vary).

Our assumptions are appropriate for the setting we study. The goals that national security agencies pursue are things like defeating the Soviet Union in the Cold War, or winning the Second World War. In general, we believe that managers and researchers employed in the national security community benefit to the extent that they succeed in these goals and lose to the extent they fail in them. This is partly due to the extensive security clearance process and constant monitoring that national security employees are subject to. It also relates to professional incentives once in these communities. Finally, evidence suggests that public-sector employees, and especially national security employees, tend to have a strong public service motivation. However, individual agents may disagree about the best way to achieve these goals, face incentives to buck-pass, or have parochial incentives that cause them to weight the costs and benefits differently.

Studies of gambling for resurrection are closer to us because they examine a leader and the public of the same country. Most notably, Downs and Rocke (1994) theorize about the president as the agent who makes the choice to fight a war (or not). The president holds asymmetric information over whether war serves the public interest. They model the public as the principal who can re-elect the president. This model is closer to ours than the hierarchy and security force assistance literatures in that the public and the president both share a preference for avoiding bad foreign policy outcomes.

But there are several differences. First, the president has a unique incentive for re-election that can conflict with the public's. As discussed, these preferences are not appropriate in our theory

(although our theory is robust if we model preference variation like this). Second, the president has private information about the quality of the choice to fight, and his own quality. This is not appropriate in our model for two reasons. The first reason is, unlike the American public, the manager has a security clearance and access to a wide cadre of classified researchers who can review the existing data. The second reason is that the researcher is very uncertain before they engage in pilot research precisely because they have not worked on a problem like this. Third, the public directly punishes the president through an electoral mechanism. This is not appropriate in our theory for two reasons. One reason is that the manager is complicit through don't-ask-don't tell, and therefore does not do the punishing. Another is that punishment does not take the form of replacing a researcher with a different one (as in the electoral context).

References

a. Lake, David. 2001. "Beyond Anarchy: The Importance of Security Institutions." *International Security* 26:129–160.

Adamsky, Dima. 2010. *The Culture of Military Innovation: The Impact of Cultural Factors on the Revolution in Military Affairs in Russia, the US, and Israel*. Stanford: Stanford University Press.

Beard, Edumnd. 1976. *Developing the ICBM: A Study in Bureaucratic Politics*. New York: Columbia University Press.

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.

Downs, George W. and David M. Rocke. 1994. "Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes to War." *American Journal of Political Science* 38:362.

Dragu, Tiberiu and Yonatan Lupu. 2021. "Digital Authoritarianism and the Future of Human Rights." *International Organization* 75(4):991–1017.

Eisenhardt, Kathleen M. 1989. "Agency Theory: An Assessment and Review." *The Academy of Management Review* 14:57.

Farrell, Theo G. and Terry Terriff. 2002. *The sources of military change: Culture, politics, technology*. Boulder, CO: Lynne Rienner.

Garfinkel, Ben and Allan Dafoe. 2019. "How does the offense-defense balance scale?" *Journal of Strategic Studies* 42:736–763.

Gilli, Andrea and Mauro Gilli. 2019. "Why China Has Not Caught Up Yet: Military-Technological Superiority and the Limits of Imitation, Reverse Engineering, and Cyber Espionage." *International Security* 43(3):141–189.

Gohdes, Anita R. 2020. "Repression technology: Internet accessibility and state violence." *American Journal of Political Science* 64(3):488–503.

Griffin, Stuart. 2017. "Military Innovation Studies: Multidisciplinary or Lacking Discipline?" *Journal of Strategic Studies* 40(1-2):196–224.

Grissom, Adam. 2006. "The future of military innovation studies." *Journal of strategic studies* 29(5):905–934.

Hart, Oliver and Bengt Holmström. 1987. *The theory of contracts*. Cambridge University Press pp. 71–156.

Hawkins, D G, D A Lake, D L Nielson and M J Tierney. 2006. *Delegation and Agency in International Organizations*. Cambridge University Press.

Horowitz, Michael. 2010. *The diffusion of military power : causes and consequences for international politics*. Princeton University Press.

Horowitz, Michael C. 2016. "Public Opinion and the Politics of the Killer Robots Debate." *Research & Politics* 3(1):1–8.

Horowitz, Michael C., Evan Perkoski and Philip B.K. Potter. 2018. "Tactical Diversity in Militant Violence." *International Organization* 72(1):1–35.

Jensen, Benjamin M. 2016. *Forging the Sword: Doctrinal Change in the U.S. Army*. Stanford, CA: Stanford University Press.

Jungdahl, Adam M and Julia M Macdonald. 2015. "Innovation inhibitors in war: Overcoming obstacles in the pursuit of military effectiveness." *Journal of Strategic Studies* 38(4):467–499.

Kollars, Nina. 2014. "Military innovation's dialectic: Gun trucks and rapid acquisition." *Security Studies* 23(4):787–813.

Kurth Cronin, Audrey. 2020. *Power to the People: How Open Technological Innovation is Arming Tomorrow's Terrorists*. New York: Oxford University Press.

Ladwig, Walter C. 2016. "Influencing Clients in Counterinsurgency: U.S. Involvement in El Salvador's Civil War, 1979–92." *International Security* 41:99–146.

Lai, Edwin L.-C., Raymond Riezman and Ping Wang. 2009. "Outsourcing of innovation." *Economic Theory* 38:485–515.

Lee, Caitlin. 2019. "The Role of Culture in Military Innovation Studies: Lessons Learned from the US Air Force's Adoptino of the Predator Drone, 1993-1997." *Journal of Strategies Studies* pp. 1–35.

Miller, Gary J. 2005. "THE POLITICAL EVOLUTION OF PRINCIPAL-AGENT MODELS." *Annual Review of Political Science* 8:203–225.

Moghadam, Assaf. 2013. "How al Qaeda innovates." *Security Studies* 22(3):466–497.

Nielson, Daniel L. and Michael J. Tierney. 2003. "Delegation to International Organizations: Agency Theory and World Bank Environmental Reform." *International Organization* 57:241–276.

Perkoski, Evan. 2019. *Terrorist technological innovation*. Oxford: Oxford University Press.

Posen, Barry. 1984. *The sources of military doctrine: France, Britain, and Germany between the world wars*. Ithaca: Cornell University Press.

Price, John F. 2014. "US Military Innovation: Fostering Creativity in a Culture of Compliance." *Air & Space Power Journal* 43(Sep.-Oct.):128–134.

Rosen, Stephen Peter. 1988. "New ways of war: understanding military innovation." *International security* 13(1):134–168.

Sapolsky, Harvey M. 1972. *Polaris System Development: Bureaucratic and Programmatic Success in Government*. Cambridge, MA: Harvard University Press.

Sechser, Todd S., Neil Narang and Caitlin Talmadge. 2019. "Emerging technologies and strategic stability in peacetime, crisis, and war." *Journal of Strategic Studies* 42:727–735.

Stiglitz, Joseph E. 1989. *Principal and Agent*. Palgrave Macmillan UK pp. 241–253.

Zhang, Baobao, Markus Anderljung, Lauren Kahn, Noemi Dreksler, Michael C. Horowitz and Allan Dafoe. 2021. "Ethics and Governance of Artificial Intelligence: Evidence from a Survey of Machine Learning Researchers." *Journal of Artificial Intelligence Research* 71:591–666–591–666.