

# Shooting the Messenger: The Challenge of National Security

## Whistleblowing\*

Michael F. Joseph<sup>†</sup>

Michael Poznansky<sup>‡</sup>

William Spaniel<sup>§</sup>

February 8, 2021

### Abstract

Whistleblowers play an integral role in oversight. In almost every employment sector, organizational insiders who come forward to expose alleged wrongdoing are protected from retaliation. In contrast, national security whistleblowers face steep fines and jail sentences for coming forward. Why? We argue that the difficulty of verifying allegations of wrongdoing in the national security arena make it hard to condition rewards and punishments on the veracity of whistleblowers' claims. In such cases, harsh punishments prove effective for encouraging honest whistleblowing. We use mechanism design to build these claims and investigate the implications through an analysis of proposed reforms to whistleblower protection laws in the United States over the last 40 years. We also report data from elite interviews with real-world whistleblowers using interview techniques designed to test the mechanisms of formal models. This article contributes to the study of whistleblowing, disclosure dilemmas, and oversight in the covert sphere.

**Key Words:** whistleblowing, national security, oversight, transparency, mechanism design

---

\*Forthcoming, *Journal of Politics*.

<sup>†</sup>Assistant Professor, Department of Political Science, University of California, San Diego. (michael.frederick.joseph@gmail.com, [www.michaelfjoseph.com](http://www.michaelfjoseph.com)).

<sup>‡</sup>Assistant Professor, Graduate School of Public and International Affairs, University of Pittsburgh. (poznansky@pitt.edu, <http://michaelpoznansky.com>).

<sup>§</sup>Assistant Professor, Department of Political Science, University of Pittsburgh. (williamspaniel@gmail.com, <http://williamspaniel.com>).

In both the private and public sectors, whistleblowers—organizational insiders who step forward to expose alleged wrongdoing—are vital for oversight. These individuals face enormous risks because they uncover abuse, sometimes by the most powerful members of society. For this reason, most modern democracies write legislation to protect whistleblowers from reprisals and even reward them with financial inducements (Stanger, 2019; Tily, 2015, 1193–1994).

Whistleblowers from one sector are largely exempt from these protections: national security. National security whistleblowers often face steep punishments for exposing government wrongdoing (Moberly 2012, 95–96; Sagar 2013, 141–143; Vladeck, 2008, 1533–1534). In 2012, John Kiriakou was sentenced to 23 months in prison for revealing information about the U.S. torture program (Shane, 2013). One year later, Edward Snowden was charged under the Espionage Act for divulging details about mass surveillance. Outside of the United States, Katharine Gun, an employee of the United Kingdom’s signals intelligence agency, revealed an alleged scheme by the U.S. to spy on the United Nations before the Iraq War in 2003 (Norton-Taylor, 2003). She was subsequently arrested for violating the Official Secrets Act.

Why do countries legislate to fiercely punish national security whistleblowers but reward and protect those from other sectors? There are two intuitive but incomplete answers. One is that the government is in a position to abuse national security secrecy *and* write whistleblowing laws. As such, policymakers punish those who expose their bad behavior. But different branches of government perform different tasks. While executives can exploit secrecy, legislatures write laws that govern what happens to whistleblowers that come forward and how claims are verified. It is therefore unclear why legislatures do not check executive power.

Another possibility is that national security whistleblowers do not perform a public service because they typically expose classified information which damages the national interest.<sup>1</sup> Even if one accepts the premise that national security requires unique levels of secrecy, the public still wants to know when executives abuse power or display incompetence (Moberly, 2012, 118). When national security whistleblowers come forward with allegations of waste,

---

<sup>1</sup>For a summary of this argument, see Sagar (2013, 108–109).

fraud, and abuse, they are not revealing secrets the public wants to protect (Papandrea, 2014, 464–465). If anything, whistleblowers are needed in national security because secrecy makes oversight difficult (Arnold, 2019). This is especially true in the digital age when outlets that used to expose government abuse are increasingly viewed with skepticism (Popkin and Dimock, 2000). Whistleblowers may be one of the few sources the public finds credible.

We develop a model that addresses these issues. It begins from the premise that legislators face a dilemma between secrecy and oversight in how they craft national security whistleblower legislation. This dilemma stems from verification of allegations about executive wrongdoing. In a perfect world, honest whistleblowers would make accurate claims and they would be believed. But some may issue false or misleading allegations (Boot, 2018, 82). Investigating such claims and allowing the accused to present exculpatory evidence is therefore important.

In most sectors, the legislature can write laws that mandate whistleblowers' claims be thoroughly investigated and verified. In the national security sphere, however, the information executives must reveal to defend against allegations is often top-secret. Legislators are reluctant to write laws that put the executive into a position where they must reveal highly-classified information to defend against allegations of waste, fraud, and abuse. In short, national security whistleblowers are different not because we do not want them to report wrongdoing, but because the public interest seems to preclude procedures where these allegations can be verified with a high degree of confidence (Sagar, 2013, 123–124).

This has two strategic implications, which we examine using mechanism design. First, even though legislatures want national security whistleblowers with legitimate allegations of abuse to come forward, they will exempt them from protections provided to whistleblowers in other sectors. In sectors where verification is feasible, legislatures can reward whistleblowers deemed to be honest because allegations can be verified. In national security, they are unwilling to verify claims and therefore cannot condition rewards on their veracity.<sup>2</sup> If legislatures write laws to protect national security whistleblowers with unverified allegations, they create

---

<sup>2</sup>On the challenges of verifying national security claims even if the judiciary was involved and had access to relevant information, see Sagar (2013, 144–147).

perverse incentives for them to come forward with false or misleading claims.

Second, incentive structures always exist that elicit honest whistleblowers to come forward and keep dishonest whistleblowers quiet. This is true even in national security where legislatures are unwilling to verify claims. The key is that honest whistleblowers with altruistic preferences or who simply dislike the policy in question also benefit from inducing actual policy change. Whistleblowers who would make false allegations for fame, revenge, or other private benefits are deterred by the prospect of prison, professional disgrace, and fines. Since only those with credible allegations are willing to come forward and bear these costs, their claims are believable. To be clear, we do not think the legislature primarily punishes whistleblowers because they want to make their claims credible. Rather, it is a convenient byproduct that occurs because protections and rewards will not work when verification is unavailable.

We substantiate our argument using evidence about the two main actors in our model: the legislature and whistleblowers. On the legislative side, we analyze whistleblower protection reform in the United States. We show that members of Congress repeatedly introduced laws designed to increase protections for national security whistleblowers. These efforts largely failed because Congress accepted that credibly verifying national security whistleblowers' claims would require revealing classified information and that this was undesirable. Moreover, we show that Congress understood that protecting whistleblowers without credible verification would create perverse incentives for those with dishonest claims to come forward.

On the whistleblower side, we analyze new interview data with one of the most senior and prominent national security whistleblowers in American history: Thomas Drake. Using interview techniques designed to validate the mechanisms of formal models, we show that Drake publicly revealed his identity partially because he knew the costs he incurred made his claims credible. The appendix contains a brief historical analysis of Daniel Ellsberg and an interview with John Kiriakou to show that the mechanism applies in other critical cases.

Our results speak to a number of debates. To begin, they help explain how states resolve complex disclosure problems in national security (Kurizaki, 2007; Carnegie and Carson, 2019).

The conventional wisdom is that governments face a trade-off between protecting secrets and facilitating oversight (Colaresi, 2014). However, through harsh punishments, legislatures end up creating a mechanism wherein national security whistleblowers can make credible allegations of wrongdoing without revealing additional national security details. This works because verification is less important when only honest whistleblowers will come forward.

Additionally, this article has implications for executive oversight. Many scholars hold that oversight is easiest when monitoring is possible. In many sectors, this means laws that protect whistleblowers (Beim, Hirsch and Kastellec, 2014; Chassang and Padró i Miquel, 2014; Ting, 2008) or monitor the executive (Clinton, Lewis and Selin, 2014) decrease the chance of executive abuse. But the existing literature focuses on cases where it is easy to verify allegations (cf Ashworth, 2012). We show that even when it is hard, effective monitoring is still possible; this entails laws that punish those who come forward. Drawing from this insight, we provide the first game-theoretic framework to explain sector-specific variation in whistleblower laws. Others note that the national security sector is unique, but attribute the difference to the executive's stranglehold on the law-making process and a desire to keep power, or a normative desire to keep national security information, even abuses of power, secret (Moberly, 2012; Papandrea, 2014; Sagar, 2013; Stanger, 2019). Our explanation relies on the strategic consequences of verification. This article also speaks to the challenges of institutional reform in the covert sphere (Colaresi, 2014; Spaniel and Poznansky, 2018).

Finally, this article is relevant for those interested in the normative status of whistleblowing. These scholars debate the conditions under which whistleblowing is legitimate and therefore deserving of protection (Bocchiola, Ceva and Vinciguerra, 2020; Kenny, 2019; Kumar and Santoro, 2018). We identify the unique challenges associated with evaluating allegations of waste, fraud, and abuse in the national security sector and the barriers this poses to providing robust safeguards, even in cases where a whistleblower is justified (Boot, 2018) and raising issues of public interest (Ceva and Bocchiola, 2019; Kumar and Santoro, 2017).

## National Security Whistleblowing in Context

In our theory, whistleblowers act as “fire alarms” who alert the public, legislators, or other relevant entities when they observe someone—in our case, in government—abusing power (McCubbins and Schwartz, 1984). Scholars generally believe that whistleblowers benefit society because they are in a unique position to quickly expose abuse and can potentially deter it in the first place (Beim, Hirsch and Kastlelec, 2014; McCubbins and Schwartz, 1984, 166; Ting, 2008; Givati, 2016). As a result, they are often protected.

In the United States, for example, Congress passed the first significant whistleblower protections in 1978 as part of the Civil Service Reform Act. Subsequent legislation, including the Whistleblower Protection Act (1989), the Sarbanes-Oxley Act (2002), and the Whistleblower Protection Enhancement Act (2012), closed loopholes and strengthened protections (Nielsen, 2020, 672). In each case, however, Congress developed different laws to address whistleblowers from the national security sector (Bellia, 2012, 1526). While these regulations have evolved over time, there are at least two persistent ways in which national security is unique.

First, whistleblowers from most sectors can report allegations of waste, fraud, and abuse to an entity outside the agency where they work.<sup>3</sup> In the U.S., this is the Office of the Special Counsel (OSC), which can order investigations of alleged wrongdoing following a claim (Moberly, 2012, 58).<sup>4</sup> External reporting channels and investigations are important because those that abuse power face incentives not to address allegations of their own wrongdoing, and to bury reports suggesting they have abused power (Nielsen, 2020, 673–674). External investigators are also less likely to punish those who speak out (Sagar, 2013, 133).

National security whistleblowers are not permitted to report allegations to an independent actor like OSC. Instead, they must go to entities within their own agency or a designated member of the executive branch. As discussed below, recent legislative changes have tried to redress this. National security professionals can now report to select members of Congress,

---

<sup>3</sup>On internal versus external channels, see Ceva and Bocchiola (2019, 187).

<sup>4</sup>See <https://osc.gov/Pages/DOW.aspx>.

but only if the whistleblower is first authorized by an Inspector General—who is part of the executive branch. If the Inspector General does not authorize external reporting, then the employee must seek approval from the head of their agency to report externally. Otherwise, they can only use internal channels. Analysts note that these internal veto points effectively close off independent reporting options (Moberly, 2012, 108–109). Even when Congress does receive information, they may be prohibited from acting upon it (Vladeck, 2008, 1545).

A second way national security is unique turns on the laws protecting whistleblowers from employer-driven reprisals. In most sectors, those that are fired, denied promotion, or face hostile working conditions can complain to the OSC. Their case can also be adjudicated before the Merit System Review Board with the possibility of appealing to a federal court (Moberly, 2012, 129–130). National security professionals do not enjoy these protections. One of the first pieces of legislation aimed at this group—the Intelligence Community Whistleblower Protection Act of 1998—“provides no legal remedy for retaliation against a covered employee” (Papandrea, 2014, 493). In 2012, Obama issued Presidential Policy Directive 19 which allows national security whistleblowers to appeal to a panel of Inspectors General in the event of reprisals. However, the decision “is [still] subject to review by the agency head, thereby potentially mitigating much of the benefit of the outside review” (Papandrea, 2014, 496).<sup>5</sup>

All of this raises the question: Why would legislators write, and the public accept, laws that restrict robust oversight of national security? We explore this logic using a game-theoretic tool known as mechanism design. Mechanism design allows us to consider how different legislative choices influence the strategic incentives of whistleblower behavior. Before introducing the model, we will describe the incentives and strategic options available to these two actors.

### **Whistleblowers**

Existing research identifies several reasons why someone might blow the whistle (Kenny, 2019). Two of the most common are altruism and private gain (Bocchiola, Ceva and Vinciguerra, 2020). Which factors are more salient depends on the individual (Papandrea, 2014, 486).

---

<sup>5</sup>See also Sagar (2013, 118).

Altruism in this context refers to “an individual’s orientation to delivering services to people with a purpose to do good for others and society” (Caillier, 2017, 812). Some whistleblowers display high levels of pro-social attributes, which are typically associated with a public service orientation (Near and Miceli, 2008). They care less about how wrongdoing is exposed and more that it ends.

Whistleblowers may also accrue private benefits by coming forward. These include emotional benefits derived from moral narcissism (Sagar, 2013, 151), the desire for fame, or the desire to harm employers and colleagues (Near and Miceli, 1996, 509). They also include financial or professional rewards. Public sector employees may allege wrongdoing—real or perceived—in an effort to try to capitalize on legal protections that prevent them from getting fired. Whistleblowers motivated primarily by private benefits do not care if wrongdoing is actually taking place or, in the event it is, that it stops. They are motivated by the perks they accrue by alleging wrongdoing.

In addition to the incentives whistleblowers have for alleging wrongdoing, they must also consider the costs they will face for doing so. This includes formal retaliation, such as being fired, as well as informal retaliation, including social ostracization or blacklisting from employment opportunities (Near and Miceli, 2008, 267; Sagar, 2013, 6).

### **The Legislature**

We think about the legislature as the actor that establishes the rules of the game before whistleblowers decide to step forward. It has two main levers at its disposal. First, legislators can manipulate the private costs and benefits whistleblowers face. They can reward whistleblowers through safeguards against workplace reprisals such as losing one’s job. Depending on the allegation, lawmakers can provide “bounties”—monetary compensation—to whistleblowers that step forward (Tily, 2015, 1203). Legislators can also impose costs on whistleblowers, failing to protect them by withholding safeguards against employer-driven reprisals or punishing them directly by making it illegal to reveal information. As discussed above, existing laws raise the benefits for every sector except national security, where laws raise costs.

Second, legislators can manipulate how whistleblowers' allegations are vetted. Specifically, they can write laws wherein whistleblowers' claims are investigated by an independent entity or internally (Kenny, 2019, 19). They can also write laws that determine whether subsequent adjudication of an allegation, or claims of employer-driven reprisals, will be heard by entities in the executive branch or will be subject to some form of external review like a court.

## Theory

We study how oversight organizations can obtain quality information regarding violations of protocol, such as Congress discovering executive abuse. The method we examine is whistleblowing. Because oversight organizations can change the incentives whistleblowers face, we adopt an institutional design approach. From this perspective, the core challenge is whether a principle (the oversight organization) can construct an incentive structure such that its agent (the whistleblower) only reveals true and credible information. As such, we investigate the broad incentive compatibility constraints that induce a whistleblower with real grievances to come forward while convincing a whistleblower without legitimate grievances to stay silent. This may prove difficult because—staying true to the informational environment—the institutional designer can only condition incentives based on its (potentially uncertain) beliefs.

The basic set-up is not specific to the national security sector. We include a transparency parameter that allows us to moderate the ease with which the government can investigate and verify a whistleblower's claims. Setting this parameter low implies that it is difficult to investigate a whistleblower's allegations.

**The whistleblower's direct incentives.** We begin by defining the whistleblower's preferences over revelation. Let  $e(\phi, V, I) \mapsto \mathbb{R}$  model the ego payoff for the whistleblower. This incorporates an individual's feelings for altruism and zealotry for revelation. It also captures the prizes, fellowships, and book deals that may come with the fame gained as a whistleblower.

We parameterize the ego function with three inputs to guide our analysis. The first input  $\phi \in [0, 1]$  represents the legislature's and the public's posterior belief that the information

is real at the end of the game.<sup>6</sup> Meanwhile, the parameter  $V > 0$  reflects the substantive importance of the issue area. Finally,  $I$  is an indicator function that takes on 1 when the information is in the public interest and 0 when the information is irrelevant or fabricated.

We make the following assumptions about the ego function. First,  $\frac{\partial e}{\partial \phi} > 0$ ; that is, the whistleblower's utility strictly increases in the belief regarding the veracity of the information regardless of the actual truth. This reflects how, all else equal, whistleblowers prefer having others believe their claims. It also captures how legislators and the public might take greater interest in the whistleblower's actions and enhance her reputation if people perceive her information as truthful and credible.

Second, a critical value  $\phi^* < 1$  exists for which (1)  $\frac{\partial e}{\partial V} > 0$  for all  $\phi > \phi^*$  and (2)  $\frac{\partial e}{\partial V} < 0$  for all  $\phi < \phi^*$  for  $I = 0$ . This has a lot to unpack. The former condition means that, if legislators and the public are sufficiently convinced of the claim, the whistleblower's ego payoff increases in the importance of the issue area regardless of the actual truth. The latter condition implies that if they are sufficiently skeptical of the claim, the ego payoff for a false claim decreases in the importance of the issue area. For example, holding fixed complete disbelief, a low-value accusation will receive less negative attention than a high-value accusation. Thus, the whistleblower internalizes more embarrassment in the latter case.

Finally, let  $e(\phi, V, 1) > e(\phi, V, 0)$  holding fixed any values of  $\phi$  and  $V$ . That is, keeping beliefs and the importance of the issue area constant, revealing public-interest information gives the whistleblower a greater payoff than revealing unimportant or fabricated information. This captures the pro-social motives of altruistic whistleblowers who benefit from stopping waste, fraud, and abuse. It also incorporates a morality payoff for doing right thing by stepping forward (Alford, 2001).

**The institutional designer's toolkit.** The institutional designer can manipulate the whistleblower's incentives in two ways. First, it selects a function  $c(\phi) \mapsto \mathbb{R}$  that maps a posterior belief to a cost that the whistleblower receives regardless of all other factors. This

---

<sup>6</sup>This is not the whistleblower's belief, as the whistleblower is fully informed.

represents the hardship that legislation imposes on whistleblowers. Greater protections for whistleblowers imply a lower  $c$  value. Because  $c$  is a function of  $\phi$ , it captures any financial incentives a whistleblower receives based on verified information; these manifest as “negative” costs. We restrict attention to functions where  $c'(\phi) \leq 0$ , meaning that punishments become weakly lighter as the belief that the information is genuine increases. Because the inequality is weak, it allows for incentives and punishments to stay constant in the belief. The designer could, for example, choose for trivial functions like  $c(\phi) = c$ .

The second tool the designer has is investigative transparency, a process that can verify the whistleblower’s claim. Let  $t \geq 0$  represent the government’s investigative transparency. Investigative transparency determines two parts of the incentive structure. First, let  $p_T(t) \mapsto [0, 1)$  be an exogenous function that maps a level of investigative transparency to the probability that Nature reveals the accusations are true following a genuine whistleblower coming forward. Similarly, define  $p_F(t) \mapsto [0, 1)$  as an exogenous function that maps a level of investigative transparency to the probability that Nature reveals the accusations are false following an insincere revelation.<sup>7</sup> Let  $p_T(0) = p_F(0) = 0$ , meaning that an institution with no investigative transparency implies no revelation. We also assume that  $\frac{dp_T}{dt}, \frac{dp_F}{dt} \geq 0$ , so that transparency does not make the truth more opaque. The weak inequality allows us to investigate especially difficult issue areas where investigative transparency has no effect on verification.<sup>8</sup>

**The legislator’s challenge.** Putting everything together, the overall payoff for a genuine whistleblower coming forward is:<sup>9</sup>

---

<sup>7</sup>Allowing  $p_T$  or  $p_F$  to equal 1 does not change our main results but adds to the number of separate cases we must analyze. This is because, if the designer chooses a  $t$  value to make one of the functions equal to 1 but not the other, the other is perfectly informative by implication. For example, suppose  $p_T = 1$ . Then the only way an investigation can fail to reveal the nature of the accusation is if the accusation was disingenuous. Because of that transparency, it is easier to build the desired incentive structure than in the cases we address.

<sup>8</sup>To be clear, we allow the designer to legislate different levels of investigative transparency. Our point is that we explicitly assume setting weak transparency makes verification difficult.

<sup>9</sup>The first term substitutes  $\phi = 1$  because of the player’s type. Similarly, the first term in an insincere whistleblower’s payoff substitutes  $\phi = 0$ .

$$\underbrace{p_T(t)}_{\text{Pr. Revelation}} \overbrace{(e(1, V, 1) - c(1))}^{\text{Value: Honest Revelation}} + \underbrace{(1 - p_T(t))}_{\text{Pr. No Revelation}} \overbrace{(e(\phi, V, 1) - c(\phi))}^{\text{Value: Honest No Revelation}}$$

Meanwhile, the overall payoff for an insincere whistleblower coming forward is:

$$\underbrace{p_F(t)}_{\text{Pr. Revelation}} \overbrace{(e(0, V, 0) - c(0))}^{\text{Value: Dishonest Revelation}} + \underbrace{(1 - p_F(t))}_{\text{Pr. No Revelation}} \overbrace{(e(\phi, V, 0) - c(\phi))}^{\text{Value: Dishonest No Revelation}}$$

Finally, we standardize both types' payoffs for staying silent to 0.

We want to know the following. Suppose the government only has a prior belief  $q \in (0, 1)$  that a violation has occurred. Can the government choose a disutility function  $c(\phi)$  and an investigative transparency  $t$  such that the genuine whistleblower wishes to come forward and the illegitimate whistleblower wishes not to? If so, what are the trade-offs associated with obtaining credible information and suppressing improper revelation?

### Can Laws Elicit Credible Whistleblowing?

On the existence of effective whistleblowing mechanisms, we begin with some good news:

**Proposition 1.** *For all parameters, there exists a value of  $t$  and a function  $c(\phi)$  such that genuine whistleblowers reveal information and disingenuous whistleblowers do not.*

Proposition 1 is encouraging. It implies that legislatures can, in principle, write a set of laws to guarantee that whistleblowers who observe wrongdoing will come forward and report abuse and that no one will come forward with a false claim. If the legislature passes these well-calibrated laws, members of civil society can be confident that any whistleblower allegation they observe is genuine.

Demonstrating the claim requires checking the incentive compatibility constraints. First, consider the genuine whistleblower. She is willing to come forward if:

$$p_T(t)(e(1, V, 1) - c(1)) + (1 - p_T(t))(e(\phi, V, 1) - c(\phi)) \geq 0$$

If the types separate as desired, the observer can perfectly update their information regardless of Nature's revelation. This is because, for these separating strategies, only genuine whistleblowers come forward. Thus, even if Nature does not verify the accuracy of the claim, the whistleblower must have genuine public-interest information because the disingenuous whistleblower remains silent. This implies that  $\phi = 1$ . Making that substitution generates:

$$p_T(t)(e(1, V, 1) - c(1)) + (1 - p_T(t))(e(1, V, 1) - c(1)) \geq 0$$

$$c(1) \leq e(1, V, 1) \tag{1}$$

Now consider the disingenuous whistleblower. Keeping her silent requires:

$$p_F(t)(e(0, V, 0) - c(0)) + (1 - p_F(t))(e(\phi, V, 0) - c(\phi)) \leq 0$$

We can again rewrite the second term. In a separating equilibrium, if the whistleblower without real information were to deviate, she would be fully believed. Substituting  $\phi = 1$  once more yields:

$$p_F(t)(e(0, V, 0) - c(0)) + (1 - p_F(t))(e(1, V, 0) - c(1)) \leq 0 \tag{2}$$

If disingenuous information is always revealed such that  $p_F(t) = 1$ , this reduces to:

$$e(0, V, 0) \leq c(0) \tag{3}$$

Otherwise, the less verifiable cases require that:

$$c(1) \geq \frac{p_F(t)(e(0, V, 0) - c(0))}{1 - p_F(t)} + e(1, V, 0) \tag{4}$$

To obtain the desired goal, the designer simply needs to choose values for  $t$ ,  $c(0)$ , and  $c(1)$  that simultaneously satisfy Lines 1 and 3/4.<sup>10</sup>

Given so much control over the situation, it might not be surprising that the designer can elicit separation under some circumstances. For example, suppose that investigative transparency was very effective, such that the government could choose a  $t$  value that set

---

<sup>10</sup>Thus, the disincentive function  $c(\phi)$  necessary to obtain separation boils down to picking just two values:  $c(0)$  and  $c(1)$ .

$p_F(t) = 1$ . Then it is easy to elicit the desired response. There could be no disutility associated with genuine information, and the disutility for disingenuous information just needs to be larger than the value for coming forward.

However, Proposition 1 states that the claim holds for all parameters. This includes circumstances that seem stacked against genuine revelation. For example, suppose that  $p_F(t) = 0$  for all  $t$ —that is, no matter the reforms to investigative transparency, no amount of verification is possible. Then we can rewrite Line 4 as:

$$c(1) \geq e(1, V, 0) \tag{5}$$

Stringing together Lines 1 and 5, we can still obtain the desired outcome by choosing a  $c(1)$  such that:

$$c(1) \in [e(1, V, 0), e(1, V, 1)] \tag{6}$$

For an institutional designer to have the capability to choose such a  $c(1)$ , the lower bound must fall below the upper bound. Thus, a mechanism designer can induce separation for all  $p_F(t)$  if  $e(0, V, 0) \leq e(1, V, 1)$  and  $e(1, V, 0) \leq e(1, V, 1)$ . The latter is true by the earlier assumption that whistleblowers earn a greater benefit from revealing a genuine infraction than by making a disingenuous claim, all else equal. The former is true because  $e(0, V, 0) < e(1, V, 0)$  due to the assumption that the whistleblower’s utility strictly increases in the belief that the information is true. As such, values for  $c(1)$  exist that fulfill the requirements.

To provide better context, the designer’s goal is to make the disutility large enough to confer credibility while not making it so large to dissuade genuine whistleblowers. Where the designer must place  $c(1)$  depends on the characteristics of the function  $e$ . For example, if an individual’s zealotry shifts both  $e(0, V, 0)$  and  $e(1, V, 0)$  higher, the designer may have to raise  $c(1)$  accordingly. But the primitives of the model guarantee such a  $c(1)$  value will still exist.

One might wonder what happens if multiple government entities had different beliefs about a whistleblower’s credibility before the whistleblower came forward. In fact, they would not

vary in how they interpret the allegation. Indeed, as the derivation of the conditions showed, the prior belief does not change the construction of  $t$  or  $c(\phi)$ .

The key here is that the designer wants to induce separation. Suppose the designer chooses values of  $t$ ,  $c(0)$ , and  $c(1)$  that satisfy Lines 1 and 3/4. Under the corresponding separating equilibrium, an observer anticipates that the honest type reveals and the disingenuous type does not. The observer's *ex ante* belief does not play into this inference. Indeed, a disingenuous type could make an accusation *and would be taken at her word*. However, the design of  $t$ ,  $c(0)$ , and  $c(1)$  is such that lying in this manner is not in the disingenuous type's best interest. As a result, regardless of one's initial belief, the separating mechanism works.<sup>11</sup>

### Trade-offs

Proposition 1 shows that legislatures do not need external investigations of whistleblower claims to ensure only honest whistleblowers come forward. This is good news for national security legislators who are, as we argued above in our substantive discussion, generally unwilling to force the executive to reveal national secrets to defend herself against allegations. But Proposition 1 does not explain how certain laws work together or against one another to achieve the desired goal. We now examine how the institution must adjust costs and benefits in a world where verification is difficult:

**Proposition 2.** *Suppose that greater investigative transparency reveals more disingenuous information (i.e.,  $\frac{dp_F}{dt} > 0$ ) and the mechanism designer lowers investigative transparency levels. Then it must weakly increase the disincentives (i.e.,  $c(0)$  or  $c(1)$ ) to maintain the disingenuous type's compliance.*

Proposition 2 reveals a nefarious trade-off on how legislatures must treat whistleblowers to obtain credible claims in a world where they are unwilling to investigate claims ex-post: they must raise the costs whistleblowers incur for coming forward with an allegation.

---

<sup>11</sup>One way that disparate beliefs would begin to matter is if some portion of the audience were "conspiracy theorists" that believe any accusation is true regardless of any rational updating process. This would alter the  $\phi$  input of the  $e(\phi, V, I)$  function. Accordingly, the condition from Line 3/4 fluctuates based on the prevalence of those conspiracy theorists.

We can observe this by examining Line 2. The derivative of the function with respect to  $t$  is  $\left(\frac{dp_F}{dt}\right)(e(0, V, 0) - c(0) - (e(1, V, 0) - c(1)))$ . Because  $e(0, V, 0) < e(1, V, 0)$  and  $c(1) \leq c(0)$ , the function decreases in  $t$ . The previous subsection showed that a range of values permits separation. Thus, one of two things is true about decreasing investigative transparency. If the decrease is marginal and the disingenuous type's condition was not binding, then separation still works without any further changes. But if the increase is sizable or the condition was binding, then the designer must do something to compensate. The two levers it has at its disposal are altering the disincentives  $c(0)$  and  $c(1)$ . Both decrease the disingenuous type's utility and thus work to offset the problem. Note further that if there were some limits to punishment and the designer insists on low investigative transparency,  $c(1)$  must remain high. This is because  $c(0)$  only has so much effect when Nature rarely reveals the disingenuous claim.<sup>12</sup> Perversely, keeping disincentives for genuine whistleblowers high keeps disingenuous whistleblowers in line.

Unfortunately, this trade-off is amplified in highly salient political issues such as national security. Our model operationalizes the value of that issue area with  $V$ . We observe that the designer must be careful with low investigative transparency and high stakes:

**Proposition 3.** *Suppose the mechanism designer wishes to implement sufficiently low investigative transparency. Then as the saliency of the issue area increases, the mechanism designer must weakly increase the disincentives (i.e.,  $c(0)$  or  $c(1)$ ) to maintain the disingenuous type's compliance.*

Again, the proof for this comes from examining Line 2. The derivative of the function with respect to  $V$  is  $p_F(t) \left(\frac{\partial}{\partial V} e(0, V, 0)\right) + (1 - p_F(t)) \left(\frac{\partial}{\partial V} e(1, V, 0)\right)$ . The former partial derivative is negative—when no one believes the whistleblower, increasing the saliency of the issue area causes more embarrassment. The latter partial derivative is positive—when everyone believes the whistleblower, increasing the saliency of the issue area bestows greater pride. The derivative of the function is a convex combination of the two, determined by  $t$ . Thus,

---

<sup>12</sup>In fact, selecting a  $t$  such that  $p_F(t) = 0$  requires  $c(1)$  to increase.

sufficiently low investigative transparency suppresses  $p_F$ , thereby making the total derivative positive. If the change is large enough, Line 2 no longer holds, and so the disingenuous type fails to comply. The designer must do something to compensate, and increasing the disutility for coming forward serves that purpose.

However, note that Proposition 3 indicates that this logic does not apply globally. Although sufficiently low values of  $p_F$  guarantee the above incentives, it is possible the institutional designer can *reduce* disincentives as the issue area becomes more important if  $p_F$  is high. Imagine that  $e(0, V, 0)$  is negative. That is, the whistleblower suffers embarrassment when others know she is disingenuous. Increasing the value of the issue area exacerbates that embarrassment. Given that, suppose investigative transparency is high and will therefore often reveal disingenuous accusations. The extra embarrassment for making a claim overrides the additional praise she could earn through deception. Consequently, the institutional designer could reduce  $c(0)$  or  $c(1)$  under these circumstances and still induce honest revelations.

### Limits

Before continuing, it is worth noting the limits on the generality of Proposition 1's claims. We highlight three key components that drive the results:

**Information.** As constructed, the model assumes that the designer knows the whistleblower's preferences and can therefore perfectly calibrate the costs. In practice, whistleblowers vary in their personal incentives. One might wonder if the designer can still create a mechanism that separates honest and disingenuous whistleblowers.

In fact, because Line 6 holds for a range of values, this result is robust to that type of uncertainty, provided it is not too severe. For example, suppose that the whistleblower's payoff for coming forward has a stochastic shock of  $\epsilon \in [-\frac{\hat{\epsilon}}{2}, \frac{\hat{\epsilon}}{2}]$  that only the whistleblower observes. This might represent an individual's "zealotry" for coming forward on the subject regardless of how others will interpret it. Obtaining the incentive compatibility requirements becomes harder because the mechanism must still induce the most negatively shocked gen-

uine whistleblower to come forward while keeping the most positively shocked disingenuous whistleblower silent. Reworking Lines 1 and 5 accordingly, the equivalent to Line 6 is now  $c(1) \in [e(1, V, 0) + \frac{\hat{\epsilon}}{2}, e(1, V, 1) - \frac{\hat{\epsilon}}{2}]$ . Thus, values of  $c(1)$  still exist if  $\hat{\epsilon}$  is less than the difference between  $e(1, V, 1)$  and  $e(1, V, 0)$ . In other words, the government can still resolve the problem as long as the unobserved shock to the whistleblower's payoff is not too large.<sup>13</sup> Moreover, this is easier to fulfill as  $V$  increases if one assumes that, as the issue area becomes more important in an environment where everyone believes all claims, a genuine whistleblower's benefit increases at a faster rate than a disingenuous whistleblower's. This is intuitive—both share the same direct ego benefit, but the genuine whistleblower also causes real policy change.

**Tailoring.** When designing the institution, the model looks at a snapshot of two types of the same individual, one with genuine information and one with disingenuous information, and then builds an incentive structure around those potential types' preferences. This type of tailoring is important. Suppose, for example, an institutional designer created a single law that created a punishment schedule that took only a posterior belief and a value of the issue area as its inputs. Such a design may elicit separation for some types of whistleblowers but not others. Imagine that the cost set was substantial jail time, and consider the case of an individual without familial commitments. Then a genuine type might be willing to come forward while a disingenuous type might be deterred. In contrast, consider a case of an individual with familial commitments. Substantial jail time would likely deter both types.

Tailored punishments solve this problem. The designer ought to consider not just posterior beliefs and values of the issue area when selecting costs. Indeed, the critical element is to consider how a whistleblower incorporates those factors into her utility function and then output the appropriate punishment levels. More formally,  $\phi$  and  $V$  are not what the designer inputs into the institutional rules. Instead, it is  $e(\phi, V, 0)$  and  $e(\phi, V, 1)$ . Absent that,

---

<sup>13</sup>If it is too large, then we would have an impossibility result. The government would then choose whether to design an institution that only elicits genuine information but does not guarantee that all whistleblowers with genuine information come forward, or an institution that elicits all whistleblowers with genuine information to come forward but also mixes in some with disingenuous information.

separation requires that a particular individual just so happens to fall within the designed constraints.

**Flexibility.** In the model, the designer has control over the disutility function.<sup>14</sup> In practice, government entities may face legal or practical problems that restrict their choices. If such limits prevent the designer from choosing a  $c(1)$  that satisfies Line 6, then Proposition 1 fails. But even if just one of the permissible values fits that constraint, the result still holds.

Substantively, two problems can arise here, both related to needing to make  $c(1)$  at least  $e(1, V, 0)$ . First, constitutional limitations on punishments may restrict the designer's choice domain. For example, if the death penalty is necessary to offset the incentive to expose misleading information, constitutional provisions may prohibit implementing such laws. A second, more layered problem arises with how the broader public may react to a punishment. If giving a larger punishment itself confers a counter-response from the public, then raising the actual disutility may not be possible. For example, even if the death penalty were legal, using it may martyr the whistleblower, which itself creates an incentive to come forward.

A related problem arises from a behavioral perspective. Suppose some members of the public do not work through the separating logic that underlies the designer's decision. Then they may see a high cost as automatically conferring credibility. Accordingly, the designer's ability to raise the disutility to the necessary level may not be possible.

## **Evidence from the United States**

We validate our theory using qualitative evidence from whistleblower reform efforts and behavior in the United States. While many modern democracies have whistleblower laws that exempt national security professionals (Boot, 2018, 71-72), the U.S., which has the largest national security apparatus of any modern democracy, is an important case.<sup>15</sup> In this vein, much of the legal and policy writing on whistleblowers focus on it (Mistry and Gurman, 2020).

---

<sup>14</sup>The only restriction we placed was that the disutility weakly decreases in the belief.

<sup>15</sup>For a sample of national security secrecy laws in democracies, see Colaresi (2014, 67-68). For a discussion of whistleblowing in the European context, see Bocchiola, Ceva and Vinciguerra (2020).

In what follows, we report evidence on our two main actors. First, we analyze debates in the United States surrounding whistleblower legislation over the last 40 years. We start with the establishment of the first real whistleblower protection laws from 1978 which explicitly exempted national security professionals and then investigate the repeated, and largely failed, reform efforts to extend robust protections to them in the ensuing years. Second, we analyze in detail the case of Thomas Drake—a whistleblower from the National Security Agency (NSA)—using a semi-structured, in-person interview.

### **Predictions About the Designer**

Our model describes an inescapable trade-off between laws requiring an open investigation into whistleblower allegations, and how whistleblowers must be punished just for coming forward under the assumption that Congress wants to elicit allegations they can rely on. This leads to three predictions about how Congress should debate whistleblower reform in the national security sector. First, we should not observe laws that reward whistleblowers for coming forward and also conceal investigations into their claims. In sectors where Congress rewards whistleblowers, they should also set procedures for thorough investigation of allegations.

Second, when calls for more protections for national security whistleblowers arise, the debate should focus on the trade-off we identify. If we are right, Congress will recognize that there is a risk of dishonest whistleblowers if allegations cannot be rigorously vetted. However, credible, external investigations require the executive to lay bare sensitive information to vindicate themselves. Those who lobby against reform should emphasize that this trade-off cannot be broken. They should also argue that any system that required the executive to expose state secrets during an investigation would not be worth any benefit that could come from oversight. Critically, these arguments will emphasize the consequences of investigating whistleblower claims, and not emphasize the government’s desire for total secrecy in national security. Those who lobby for reform should argue that the trade-off is not as severe as the government believes and that it is therefore possible to have credible investigations that do not jeopardize sensitive information. They should also argue that the oversight that is

produced outweighs any cost from disclosures during the investigation. These arguments will emphasize the consequences arising during investigations about whistleblower claims and not the public's inherent right to know about abuse or the absolute need to end it.

Of course, the laws that policymakers eventually implement depend on how they prioritize protecting whistleblowers versus opening up the investigative process. When policymakers find in favor of laws that punish national security whistleblowers, they should justify it based on their priority for keeping the investigative process secret.<sup>16</sup>

Finally, since policymakers want to stop waste, fraud, and abuse, they will try hard to protect whistleblowers who make allegations in their own agency. Policymakers will also try to force the executive to act on allegations and end abuse. The extent to which lawmakers can force the executive to self-police will determine the applicability of the trade-off we identify. If Congress can do this easily, our trade-off can be resolved without resorting to punishments. If Congress repeatedly tries and fails to develop safe internal channels for whistleblowers, the trade-off we identify is a persistent feature of national security legislation.

### **The Designer: Whistleblower Reform in the United States**

Congress has long believed that national security whistleblowers do the nation an important service at great personal risk. Senator Daniel Akaka made this clear when he introduced the Whistleblower Protection Enhancement Act (WPEA) to Congress in 2009. He noted that whistleblowers play a crucial role in “alerting Congress and the public to government wrongdoing and mismanagement, protecting our civil rights and civil liberties, helping to keep us safe, and rooting out waste, fraud, and abuse.” Nevertheless, he pointed out that “[f]or too long, national security whistleblowers have not had secure avenues to disclose government waste, fraud, abuse, and mismanagement. Some undoubtedly have stayed quiet, while some have leaked classified information to the media. We must ensure that there are secure channels to bring problems in the Federal government to Congress’s attention” (Akaka, 2009). The legislation that eventually passed did no such thing. Our theoretical model explains why.

---

<sup>16</sup>This is different from the intrinsic need to keep all national security behavior secret.

It is useful to begin by analyzing the deliberations surrounding the Civil Service Reform Act of 1978. Both political parties at the time supported new protections for government employees who exposed waste, fraud, and abuse. Yet, some officials worried that these could encourage false or misleading allegations. Alan Campbell, the head of the U.S. Civil Service Commission, wrote a memo to Senator Abraham Ribicoff who had sponsored the legislation, warning: “Overbroad whistle-blower protection measures enhance existing incentives to use ‘leaks’ as a bureaucratic political weapon against policies or practices with which an employee simply disagrees. This prospect is a very serious concern and the sham whistle-blower tactic has already been used in at least one major instance” (Campbell, 1978, 7).

The challenge was how to craft legislation in such a way that whistleblowers who serve a legitimate oversight function would be protected without undermining the entire system by flooding it with false reports of wrongdoing. Campbell’s description of the mechanics of the CSRA are illuminating: “The Administration believes that the gap in its original whistleblower protection proposal can be filled, without exacerbating the problems sketched above,” such as the prospect of so-called sham whistleblowers. The solution, built into the legislation, was the creation of “an entity which has the independence and capability to investigate, apply remedies, and report to Congress and the public” thereby allaying these and other concerns (Campbell, 1978, 7). While the CSRA eventually passed overwhelmingly in the House and Senate, national security employees were explicitly excluded from it.

Consistent with our theory, the rationale for this was based on challenges in investigating these allegations rather than wholesale opposition to national security whistleblowers coming forward. As then-CIA Director Stansfield Turner put it in a speech at the National Press Club, “Many in the Intelligence Community have come to recognize the positive values for us in the oversight process. Ultimate accountability is essential to responsible action.” But, he said, “I hope that you will recognize that when we balk, for instance, at disclosing all the secrets necessary to prosecute a case in court, we do not do so in an arbitrary manner” (Turner, 1978, 10). When the CSRA was being considered in 1978, a classified memo to Turner stated: “This

legislation is unacceptable” in part “because it would ... provide inadequate and insufficient protection for intelligence sources and methods... and result in the judicial review of any intelligence information involved in Agency personnel actions” (Acting Legislative Counsel, 1978, 2). Deputy Director of the CIA Frank Carlucci similarly argued at the time that the intelligence community should be exempted from the Special Counsel of the newly-created Merit Systems Protection Board “because [they] would be given extraordinary powers to investigate allegations of reprisals against whistleblowers” (Carlucci, 1978, 2).

What these arguments boil down to is that the oversight mechanism Campbell described, which was necessary to prevent dishonest whistleblowers from exploiting the system for protections, would force the executive to share an enormous amount of evidence during investigations. Policymakers argued that the risk of exposing national security secrets in an open investigation was not worth the potential benefit. Congress could not design a different mechanism to verify national security whistleblowers’ claims and, ultimately, exempted them.

Policymakers have repeatedly tried to extend protections to national security whistleblowers since 1978. One recent attempt was the WPEA which sought to enhance protections across the board. Congress remained aware of the perverse effects of protecting whistleblowers without independent vetting. Congressman Elijah Cummings stated: “One of the arguments that opponents of expanded whistleblower protection [make] is that [it] will give a forum to people who just want to complain about management, or worse, are vindictive against their employer and want to get even” (House Oversight and Government Reform Committee, 2009).<sup>17</sup>

During hearings on the WPEA in 2009, Congress brought in diverse stakeholders to better understand the issues involved. Deputy Assistant Attorney General Rajesh De represented the White House (Stanger, 2019, 117). As our theory predicts, De did not argue that the executive had a right to absolute secrecy in matters of national security. Rather, his argument centered on the implications of subjecting claims of reprisals by prospective whistleblowers

---

<sup>17</sup>Members of the executive branch expressed similar concerns. When the initial version of the WPEA was introduced, for example, the White House held that “H.R. 985 would permit an employee to make an individualized determination — without further review and perhaps without all relevant information — to disclose classified information” (Executive Office of the President, 2007).

to an external judicial process: “[W]ith respect to national security whistleblowers, we think district court review and jury trials is particularly inappropriate in that context given the sensitive nature of the information at issue and the potential for wide-ranging disclosure in district court” (Subcommittee on Oversight of Government Management, 2009).

De’s remarks echoed the White House’s position from two years earlier. A 2007 memo argued that any law providing for external investigations of whistleblowers’ allegations of wrongdoing and any subsequent reprisals was unacceptable. Independent adjudication of such matters, they held, would grant entities outside the executive branch the ability to review an enormous amount of highly classified information and present it to the merit board or in court. This “essentially would require the agency to choose between protecting national security information in court or conceding lawsuits.” Because the executive was unwilling to resolve such matters openly, it “would require that the matter at issue be resolved in favor of the plaintiff” since the government would necessarily have to invoke the state secrets privilege, denying themselves a meaningful defense (Executive Office of the President, 2007).

On the other side of the hearing over the WPEA in 2009 was Danielle Brian from the Project on Government Oversight. Brian made two arguments for enhanced protections, both of which support our expectations. First, she argued that the cost of closing off open investigations was larger than Congress realized. Brian reasoned that “[i]t is in the self-interest of the Congress, perhaps most importantly, to encourage those who are aware of wrongdoing to make their disclosures to Congress. Formal briefings from agency heads have their place, but they do not truly inform the Congress of the real goings on at an agency.” Brian went on to argue that “[b]y not providing real protections for national security whistleblowers, we are actually driving them to the press and encouraging leaks of classified information. That is a lose-lose situation.” In effect, she held that even if the trade-off we identify above exists, Congress should prioritize investigations given the high costs of closed-door investigations.

Moreover, Brian argued that legal innovations made it possible to conduct independent investigations that would remain secret. This extended to jury trials for national security

whistleblowers: “[T]here should be no concerns about providing intelligence agency employees with full due process rights including jury trials given that the courts already have a long history of ... handling classified materials and knowing how to manage those problems” (Subcommittee on Oversight of Government Management, 2009). In this line of reasoning, Brian is essentially positing that it is possible to break the trade-off we identified.<sup>18</sup>

In the end, Congress sided with the White House. A report from the Committee on Homeland Security and Governmental Affairs in 2012, the year the WPEA officially became law, directly referenced De’s earlier testimony. It noted that while “the Committee believes that the provisions on classified information contained in previous versions of the legislation are consistent with Congress’s constitutional role, the Committee in the 111th Congress accommodated the Administration’s concerns by adopting a compromise provision and the sponsors of the legislation in the 112th Congress included that compromise provision in [the WPEA]” (United States Senate, 2012). In short, national security was exempt from this legislation.

Before closing, it is worth noting that while Congress excluded most of the national security sector from the WPEA, they still wanted whistleblowers to play an oversight role. As part of the 2014 Intelligence Authorization Act, they built on the earlier Intelligence Community Whistleblower Protection Act of 1998 by codifying parts of PPD-19 which strengthened internal reporting mechanisms for those in the intelligence community who expose waste, fraud, and abuse. (DeVine, 2019). Analysts have noted the limitations of these provisions (Papandrea, 2014; Stanger, 2019). But the fact that Congress repeatedly tries and fails to meaningfully bolster self-policing highlights the reality that while they attempt to circumvent the trade-offs associated with external reporting (e.g. by enhancing internal reporting), they have so far been largely unsuccessful.

---

<sup>18</sup>We do not think this is the case. But we reason that anyone that wanted to convince Congress to raise protections would try hard to convince Congress that they could keep state secrets secret in an open investigation.

### **Predictions About the Whistleblower**

Given these legislative conditions, we might expect that national security professionals never come forward publicly and abuses largely go unnoticed. After all, whistleblowers can report their concerns to the Inspector General's office in the agency that employs them at little to no cost. If they go public, they face jail time or worse. Nevertheless, many whistleblowers have come forward publicly and faced criminal charges from the government. Why?

One possibility is that the laws are not calibrated correctly. Another is that whistleblowers believe the public has an inherent right to know about waste, fraud, and abuse (Halperin and Hoffman, 1976). We suggest a more nuanced interaction. When the costs of making public accusations are high, disingenuous whistleblowers stay silent because they view the risk of prosecution as sufficiently large relative to the benefits of publicity. In contrast, genuine whistleblowers interested in stopping waste, fraud, and abuse are willing to endure those costs. Recognizing that those costs deter disingenuous whistleblowers, genuine whistleblowers can calculate that their accusations will be believed, which may further reinforce the decision to come forward.

The whistleblower's desire to effect change helps explain why we sometimes observe them going public rather than using internal channels. Of course, whistleblowers generally want to avoid costs where possible and so responsive internal channels are appealing. However, those responsible for the internal channels are appointed by the executive and therefore may have similar beliefs and legal interpretations on core national security questions as they do. In such cases, internal reports are unlikely to end abuse. Going public may be the only solution.

### **The Whistleblower: The Case of Thomas Drake**

We verify our mechanism by process tracing a single case in detail: Thomas Drake's whistleblowing at the NSA. Drake's decades of experience in the military and intelligence community meant that he was readily able to understand the consequences of his choices.

Previous research provides relevant background. In the 1990s, the NSA started to develop

an automated surveillance program called Thinthread (Bamford, 2008, 44–47; Hillebrand, 2012, 698). Although it could not discriminate between American citizens and foreign nationals at the collection stage, it could encrypt Americans’ communications until a warrant was provided (Stanger, 2019, 126–127). In 2000, the NSA ceased its support for Thinthread. Automated surveillance became important again after 9/11 (Priest and Arkin, 2011). Instead of using Thinthread, however, NSA Director Michael Hayden authorized the development of a different program known as Trailblazer. It was controversial because “[i]t entirely lacked the privacy protections that were built into Thinthread,” was run by private contractors, and cost significantly more (Stanger, 2019, 127–128). The White House also initiated a program “authoriz[ing] warrantless collection of metadata on U.S. phone calls” (Stanger, 2019, 130).

Our theory makes unique predictions about the costs and benefits whistleblowers like Drake anticipate before coming forward, as well as their motivations for going public rather than using internal channels. Existing accounts about Drake do not systematically address these dynamics. To get this information, we interviewed him in July 2019.<sup>19</sup> We contacted Drake using the solicitation technique recommended by Peabody et al. (1990). We then used semi-structured techniques recommended for elites interviews that validate causal mechanisms (Aberbach and Rockman, 2002; Harvey, 2011). Among other things, these techniques help address response bias and social desirability bias and ensure that the hypotheses can be falsified. The appendix elaborates on our design choices and illustrates our mechanism in two other important cases: a brief historical analysis of Daniel Ellsberg and the Pentagon Papers and an interview with John Kiriakou who blew the whistle on torture in 2007.<sup>20</sup>

Our interview with Drake focused on two broad questions to test our mechanism: (1) Why did Drake start with internal reporting and then switch to external reporting? and (2) Why did Drake reveal his identity to reporters rather than provide anonymous information?

---

<sup>19</sup>A third component of Drake’s whistleblowing related to what he saw as the NSA’s failure to share relevant information that might have prevented the 9/11 attacks.

<sup>20</sup>We repeated the same process in our elite interview with Kiriakou as we did with Drake.

## Internal Reporting

Starting in late-2001, Drake raised concerns about the 9/11 intelligence failure, Trailblazer, and mass surveillance internally, first with NSA leadership and then with Inspectors General at the NSA and Department of Defense. He also provided classified and unclassified details about these issues to the House and Senate Intelligence Committees (Delmas, 2015, 99).

To understand this decision, we asked Drake why he turned to internal channels first. One reason he gave supports the notion that whistleblowers are motivated in large part by public service motivations. Drake appreciated the need to maintain secrecy in the national security arena. He wanted to expose wrongdoing and have it stop but did not want to unnecessarily reveal classified information. Drake understood that the rules were put in place to protect properly classified information. In his telling, it was right to start by “follow[ing] the rules. The rules were that if you suspected wrongdoing, there were internal channels within the Agency, and then there were actually other channels that included Congress and the Department of Defense.”<sup>21</sup> Parenthetically, this helps us rule out the notion that Drake blew the whistle because he thought the public had an intrinsic right to know. To the contrary, his decision to start internally suggests that he was most interested in effecting policy change.

Another plausible reason Drake pursued internal channels before going public was timing. He was one of the first national security whistleblowers to come forward following the reforms in the late-1990s, mentioned above. As such, the new reporting mechanisms created as a result of that legislative reform had not really been tested. It may not have been obvious that they would not work. As Drake put it, the internal whistleblowing he did was done “in accordance with the Intelligence Community Whistleblower Protection Act of 1998 (ICWPA).”<sup>22</sup>

Partly as a result of Drake’s internal whistleblowing, the IG produced a report that raised serious concerns with Trailblazer, concluding that it had wasted public funds. The program was eventually shut down (Mayer, 2011). But consistent with our theory and the literature on

---

<sup>21</sup>Interview with authors, July 16, 2019.

<sup>22</sup>Correspondence with Drake, September 13, 2019.

the legal aspects of whistleblowing, these reports had little effect on the executive's broader counter-terrorism policies, including its mass surveillance program.

There are several reasons for this. First, internal investigators could not complete a thorough audit because the extraordinarily high classification markers on the parts of the programs that were of most concern to Drake. Second, the IG could not bring about change on its own (Papandrea, 2014, 473). As Drake put it, “the problem in all of this is that there’s no teeth. The IG does not have the ability to shut down programs, they can only make recommendations. That’s left up to others.” Congress was also constrained in its ability to gain access to relevant information and in terms of its enforcement capacity (Papandrea, 2014, 466–467). Moreover, effective oversight was impeded by the close relationship between Director Hayden and the Chair of the Intelligence Committee at the time, who was a strong supporter of Bush’s counter-terrorism approach.<sup>23</sup> According to Drake, Hayden would often say “he didn’t care about any staffer, all [he] care[d] about [was] the Chair, which in this case was Porter Goss. That’s the only person he cared about, because he had enormous power.”

Some national security whistleblowers since Drake have altered their behavior as a result of his experience. For example, we asked John Kiriakou why he went directly to the press rather than using internal reporting channels first. His rationale referenced Drake’s case: “I followed this case, and I thought, my god, if I go to the Oversight Committee, they’re going to charge me with espionage. If I go to the IG, they’re going to charge me with espionage.”<sup>24</sup>

### **Going to the Press**

By 2005, it was clear to Drake that internal reporting would not alert the right people to effect change. At that point he considered two options: go to the press or do nothing. “After I had done my final whistleblowing within the system,” Drake told us during the interview, “I began to consider what would it look like, knowing it would end my career and worse—worse being exactly what happened to Ellsberg. I already knew that. It was not like you were so

---

<sup>23</sup>On capture of Congress by the executive, see Sagar (2013, 134) and Stanger (2019, 168–169).

<sup>24</sup>Interview with authors, May 9, 2019.

naïve to think you could get away going to the press. I was not going to get away with going to the press. Especially in this post-9/11 world.”

As alluded to, Drake was well aware that publicity entailed significant risks: “To go to the press, it’s not just crossing the Rubicon. In fact, it’s more appropriate to say you’re going to touch the third rail. And that third rail is going to give you pretty severe shock... I knew that by going to the press that there was no question that my career would be over. And the reason for that is that if I went to the press, I’d be going in an unauthorized manner.” Worse still, he knew “they could charge [him] with espionage. [He] could end up in prison for the rest of [his] life.” The reason he was willing to expose himself to these costs was that he “wouldn’t have been able to live with [himself] for the rest of [his] life if he hadn’t done anything.” His goal was to provide information that “would be made public and the government would then have to deal with it. And it would now be out in the open, unlike all the clamps they put in place within the system and hid it.” And so, he approached the *Baltimore Sun*.

Of course, Drake could have reduced the risk of facing severe punishments by leaking information without revealing his identity. In fact, Drake first toyed with the idea of whistleblowing under cover. Initially, he reached out to reporter Siobhan Gorman anonymously to establish a secure contact in February 2006. But within a year, he dispensed with anonymity. In 2007, he walked into the Baltimore Sun and sat down on the record for several in-person interviews.

Importantly for our theory, Drake knew that exposing his identity to a reporter was extremely risky. However, he believed that doing so was necessary for two reasons. First, he said, “for me to provide the fuller context for everything I had revealed, all unclassified, to that reporter,” he said, “I would actually have to meet her.” Revealing his identity provided the necessary context, namely the fact that he was a senior-level executive at NSA and was read into the programs he was blowing the whistle on.

Second, Drake believed that exposing himself to these significant risks would raise the credibility of his story. When asked whether meeting Gorman in person increased his credi-

bility, he confirmed that it did. If he was willing to run these severe risks, it was clear that he thought the abuse was so severe that it was ultimately worth it. In his view, this was the only way to make clear to the public and Congress that something was going on that was worth investigating. Equally important, Drake acknowledged that if he opted not to go public and to remain quiet or anonymous, it would have impeded any prospect that the mass surveillance program and other abuses would end. During our interview, Drake described his decision to go to the Baltimore Sun as his “final service to the nation.”<sup>25</sup> In his view, the costs he incurred from going public, while enormous, were necessary to end wrongdoing.

## **Conclusion**

This article explored why national security whistleblowers in advanced democracies like the U.S. are punished for coming forward whereas whistleblowers in other sectors are protected and rewarded. We argued that lawmakers face a dilemma between secrecy and oversight when crafting whistleblower legislation in the national security space. Protecting whistleblowers without the ability to credibly verify allegations—which legislators have been reluctant to do—is not possible since it would create perverse incentives for individuals with false or misleading claims to come forward. While national security whistleblowers are exempt from most protections for this reason, the existence of harsh punishments renders the claims of those who do make allegations more credible. Only individuals with legitimate accusations come forward. We tested this argument by exploring the trajectory of reforms to whistleblower protection laws in the U.S. and through an interview with NSA whistleblower Thomas Drake.

This line of reasoning sheds light on several features of the recent whistleblower-driven scandal surrounding President Trump. First, many Republicans in Congress and also some members of the public did not believe the allegations (Zhou, 2019). Consistent with our logic, a major reason for skepticism is that the whistleblower did not make his or her claims in public. When pressed on why he wanted to know the whistleblower’s identity, Senator Josh

---

<sup>25</sup>Interview with authors, July 16, 2019.

Hawley (R-Mo.) stated: “How else are we going to evaluate the content and the truthfulness of these people if we don’t know who they are?” (Waldman, 2019). Second, it seems that the public impeachment trial was not an effective tool for investigating the whistleblower’s allegation. Many attribute this to partisanship. While this is important, our theory suggests that a public impeachment trial may also have structural problems that make it difficult to investigate whistleblower claims, particularly those involving issues of national security.

The research also has implications that are useful for theoretical debate and public policy analysis. Protecting whistleblowers in the national security sector is possible, but not without fundamental changes to the status quo. The existing state of affairs is such that Congress and the public accept that thoroughly vetting whistleblowers’ allegations is simply too expensive. In other words, there is a general unwillingness to require that claims be adjudicated outside the executive branch for fear that sensitive national security information could be unnecessarily compromised. Protecting whistleblowers would therefore require a radical change in how classified information is handled in such cases—namely through some kind of external review.

## **Acknowledgments**

The authors are listed alphabetically and contributed equally. They thank Peter Bills, Austin Carson, Pablo Fernandez-Vazquez, participants at the conference on national security whistleblowing held at New York University London Campus in January 2019, and the men and women of the intelligence community who informally and formally spoke to us about these issues.

## **References**

- Aberbach, Joel D. and Bert A. Rockman. 2002. “Conducting and Coding Elite Interviews.” *PS: Political Science & Politics* 35(4):673–676.
- Acting Legislative Counsel. 1978. “Civil Service Reform and Whistle-Blowing.” *CIA Electronic Reading Room (Doc. Number: CIA-RDP81-00142R000400010016-3)* .

- Akaka, Daniel. 2009. "Hearing on the Whistleblower Protection Enhancement Act." *Homeland Security and Governmental Affairs* .
- Alford, C. Fred. 2001. *Whistleblowers: Broken Lives and Organizational Power*. Ithaca: Cornell University Press.
- Arnold, Jason Ross. 2019. *Whistleblowers, Leakers, and Their Networks: From Snowden to Samizdat*. Lanham, MD: Rowman and Littlefield Publishers, Inc.
- Ashworth, Scott. 2012. "Electoral Accountability: Recent Theoretical and Empirical Work." *Annual Review of Political Science* 15(1):183–201.
- Bamford, James. 2008. *The Shadow Factory: The Ultra-secret NSA from 9/11 to the Eavesdropping on America*. New York: Anchor Books.
- Beim, Deborah, Alexander V. Hirsch and Jonathan P. Kastellec. 2014. "Whistleblowing and Compliance in the Judicial Hierarchy." *American Journal of Political Science* 58(4):904–918.
- Bellia, Patricia L. 2012. "WikiLeaks and the Institutional Framework for National Security Disclosures." *The Yale Law Journal* 121(6):1448–1527.
- Bocchiola, Michele, Emanuela Ceva and Maria Chiara Vinciguerra. 2020. Heroes or Villains? A Legislative, Ethical and Policy Assessment of Whistleblowing. In *A Research Agenda for Studies of Corruption*. Edward Elgar Publishing.
- Boot, Eric R. 2018. "No Right to Classified Public Whistleblowing." *Ratio Juris* 31(1):70–85.
- Caillier, James Gerard. 2017. "Public Service Motivation and Decisions to Report Wrongdoing in U.S. Federal Agencies: Is This Relationship Mediated by the Seriousness of the Wrongdoing." *American Review of Public Administration* 47(7):810–825.
- Campbell, Alan K. 1978. "Memo from Alan K. Campbell to Abraham Ribicoff." *CIA Electronic Reading Room (Doc. Number: CIA-RDP81-00142R000400010016-3)* .
- Carlucci, Frank C. 1978. "Memo from Frank C. Carlucci to Robert N.C. Nix." *CIA Electronic Reading Room (Doc. Number: CIA-RDP81-00142R000400010016-3)* .

- Carnegie, Allison and Austin Carson. 2019. "The Disclosure Dilemma: Nuclear Intelligence and International Organizations." *American Journal of Political Science* 63(2):269–285.
- Ceva, Emanuela and Michele Bocchiola. 2019. "Personal Trust, Public Accountability, and the Justification of Whistleblowing." *Journal of Political Philosophy* 27(2):187–206.
- Chassang, Sylvain and Gerard Padró i Miquel. 2014. "Corruption, Intimidation, and Whistle-Blowing: A Theory of Inference from Unverifiable Reports." *NBER Working Paper* w20315:1–55.
- Clinton, Joshua D., David E. Lewis and Jennifer L. Selin. 2014. "Influencing the Bureaucracy: The Irony of Congressional Oversight." *American Journal of Political Science* 58(2):387–401.
- Colaresi, Michael P. 2014. *Democracy Declassified: The Secrecy Dilemma in National Security*. Oxford: Oxford University Press.
- Delmas, Candice. 2015. "The Ethics of Government Whistleblowing." *Social Theory and Practice* 41(1):77–105.
- DeVine, Michael E. 2019. "Intelligence Community Whistleblower Protections: In Brief." *Congressional Research Service* pp. 1–10.
- Executive Office of the President. 2007. "Statement of Administration Policy: H.R. 985 Whistleblower Protection Enhancement Act of 2007." *Office of Management and Budget* .
- Givati, Yehonatan. 2016. "A Theory of Whistleblower Rewards." *The Journal of Legal Studies* 45(1):43–72.
- Halperin, Morton H. and Daniel N. Hoffman. 1976. "Secrecy and the Right to Know." *Law and Contemporary Problems* 40(3):132–165.
- Harvey, William S. 2011. "Strategies for Conducting Elite Interviews." *Qualitative Research* 11(4):431–441.
- Hillebrand, Claudia. 2012. "The Role of News Media in Intelligence Oversight." *Intelligence and National Security* 27(5):689–706.

- House Oversight and Government Reform Committee. 2009. “Rep. Edolphus Towns Holds a Hearing on H.R.1507, The Whistleblower Protection Enhancement Act of 2009.” *Congressional Quarterly* .
- Kenny, Kate. 2019. *Whistleblowing: Toward a New Theory*. Cambridge, MA: Harvard University Press.
- Kumar, Manohar and Daniele Santoro. 2017. “A Justification of Whistleblowing.” *Philosophy & Social Criticism* 43(7):669–684.
- Kumar, Manohar and Daniele Santoro. 2018. *Speaking Truth to Power. A Theory of Whistleblowing*. Cham: Springer.
- Kurizaki, Shuhei. 2007. “Efficient Secrecy: Public versus Private Threats in Crisis Diplomacy.” *American Political Science Review* 101(03):543–558.
- Mayer, Jane. 2011. “The Secret Sharer: Is Thomas Drake an Enemy of the State?” *The New Yorker* May 23.
- McCubbins, Mathew D. and Thomas Schwartz. 1984. “Congressional Oversight Overlooked: Police Patrols versus Fire Alarms.” *American Journal of Political Science* 28(1):165–179.
- Mistry, Kaeten and Hannah Gurman, eds. 2020. *Whistleblowing Nation: The History of National Security Disclosures and the Cult of State Secrecy*. New York: Columbia University Press.
- Moberly, Richard. 2012. “Whistleblowers and the Obama Presidency: The National Security Dilemma.” *Employee Rights and Employment Policy Journal* 16(1):51–141.
- Near, Janet P. and Marcia P. Miceli. 1996. “Whistle-Blowing: Myth and Reality.” *Journal of Management* 22(3):507–526.
- Near, Janet P. and Marcia P. Miceli. 2008. “Wrongdoing, Whistle-Blowing, and Retaliation in the U.S. Government: What Have Researchers Learned From the Merit Systems Protection Board (MSPB) Survey Results?” *Review of Public Personnel Administration* 28(3):263–281.
- Nielsen, Richard P. 2020. “Reformed National Security Internal Whistleblowing Systems and

- External Whistleblowing as Countervailing Ethics Methods.” *Administration and Society* 52(5):660–689.
- Norton-Taylor, Richard. 2003. “GCHQ Whistleblower Charged.” *The Guardian* Nov. 13.
- Papandrea, Mary-Rose. 2014. “Leaker Traitor Whistleblower Spy: National Security Leaks and the First Amendment.” *Boston University Law Review* 94(2):449–544.
- Peabody, Robert L., Susan Webb Hammond, Jean Torcom, Lynne B. Brown, Carolyn Thompson and Robin Kolodny. 1990. “Interviewing Political Elites.” *PS: Political Science & Politics* 23(3):451–455.
- Popkin, Samuel L. and Michael A. Dimock. 2000. Knowledge, Trust, and International Reasoning. In *Elements of Reason*. Cambridge University Press pp. 214–238.
- Priest, Dana and William M. Arkin. 2011. *Top Secret America: The Rise of the New American Security State*. New York: Little, Brown and Company.
- Sagar, Rahul. 2013. *Secrets and Leaks: The Dilemma of State Secrecy*. Princeton: Princeton University Press.
- Shane, Scott. 2013. “Ex-Officer Is First From C.I.A. to Face Prison for a Leak.” *New York Times* Jan. 5.
- Spaniel, William and Michael Poznansky. 2018. “Credible Commitment in Covert Affairs.” *American Journal of Political Science* 62(3):668–681.
- Stanger, Allison. 2019. *Whistleblowers: Honesty in America From Washington to Trump*. New Haven and London: Yale University Press.
- Subcommittee on Oversight of Government Management. 2009. “Senator Daniel K. Akaka Holds a Hearing on S.372, the Whistleblower Protection Enhancement Act of 2009.” *Congressional Quarterly* .
- Tily, Stephen Coleman. 2015. “National Security Whistleblowing vs. Dodd-Frank Whistleblowing.” *Brooklyn Law Review* 80(3):1181–1218.
- Ting, Michael M. 2008. “Whistleblowing.” *American Political Science Review* 102(2):249–267.

- Turner, Stansfield. 1978. "Protecting Secrets in a Free Society." *CIA Electronic Reading Room* (Doc. Number: CIA-RDP88-01315R000300680016-1) .
- United States Senate. 2012. *Whistleblower Protection Enhancement Act of 2012: Report of the Committee on Homeland Security and Governmental Affairs*. Washington, D.C.: U.S. Government Printing Office.
- Vladeck, Stephen I. 2008. "The Espionage Act and National Security Whistleblowing After Garcetti." *American University Law Review* 57(5):1531–1546.
- Waldman, Paul. 2019. "Sorry, Republicans. The Whistleblower's Identity No Longer Matters." *Washington Post* Nov 8.
- Zhou, Li. 2019. "Senate Republicans' Response to Allegations Against Trump: Attack the Whistleblower's Credibility." *Vox* Sep 26.

### **Biographical Statements**

Michael F. Joseph is an assistant professor in the Department of Political Science at the University of California, San Diego, San Diego, CA 92093.

Michael Poznansky is an assistant professor in the Graduate School of Public and International Affairs at the University of Pittsburgh, Pittsburgh, PA 15213.

William Spaniel is an assistant professor in the Department of Political Science at the University of Pittsburgh, Pittsburgh, PA 15213.

### Formal Appendix: Proof of Proposition 3

We wish to investigate when the lower bound of Condition 6 increases in  $V$ . Thus, we want to know the circumstances under which:

$$\frac{\partial}{\partial V} \left( p_F e(0, V, 0) + (1 - p_F) e(1, V, 0) \right) > 0$$

We can rewrite this as:

$$p_F \left( \frac{\partial}{\partial V} e(1, V, 0) - \frac{\partial}{\partial V} e(0, V, 0) \right) < \frac{\partial}{\partial V} e(1, V, 0)$$

The  $e$  function increases in  $V$  when  $\phi = 1$ . However, we make no assumptions about the rate, nor do we assume a direction for  $e(0, V, 0)$ . As such,  $\frac{\partial}{\partial V} e(1, V, 0)$  is positive but the rest is indeterminate. Consequently  $\frac{\partial}{\partial V} e(0, V, 0)$  could be greater than  $\frac{\partial}{\partial V} e(1, V, 0)$  or vice versa. In the first case, the inequality must hold because the left hand side is negative and the right hand side is positive. This corresponds to the case where increasing  $V$  requires additional punishments regardless of  $p_F$ .

If not, rearranging the inequality yields:

$$p_F < \frac{\frac{\partial}{\partial V} e(1, V, 0)}{\frac{\partial}{\partial V} e(1, V, 0) - \frac{\partial}{\partial V} e(0, V, 0)}$$

A similar case to the first can also hold if  $\frac{\partial}{\partial V} e(0, V, 0)$  is positive. This implies that the right hand side exceeds 1, thereby ensuring that it is greater than the 0-to-1 constrained  $p_F$ . Again, the institutional designer's lower bound must increase as  $V$  increases in this case.

Finally, if  $\frac{\partial}{\partial V} e(0, V, 0)$  is negative, then the right hand side forms a proper probability. If  $p_F$  is less than it, then the institutional designer must adjust  $c$  upward once more as  $V$  increases.<sup>26</sup> But if  $p_F$  is greater than it, the institutional designer can decrease  $c$  as  $V$  increases.

---

<sup>26</sup>Such a  $p_F$  must exist because the right hand side is strictly between 0 and 1 in this case.

## **Qualitative Appendix: Design for Elite Interviews**

### **Subject Selection Method**

Given the time intensive nature of case interviews and case research, and the complex nature of our predictions, we decided to focus on three whistleblowers and collect detailed information about their choices. We chose subjects based on three criteria. First, we wanted whistleblowers who made significant allegations of wrongdoing at the highest levels of government. These allegations are substantively the most important. Second, we wanted the most senior whistleblowers because these people had the most to lose and had access to the broadest set of information. Finally, we wanted whistleblowers who came forward during different eras and over different issues. Levels of trust in government varies over time, as does beliefs about the need for secrecy in government security. Furthermore, there were several legislative reform efforts that we argue had little effect. We wanted to make sure that whistleblowers at different periods of U.S. history weighed costs and benefits in the same way.

We evaluated all living whistleblowers since 1960 along these criteria. The three that fit most closely were Daniel Ellsberg, John Kiriakou, and Thomas Drake.

### **Case Analysis of Existing Materials**

Once we settled on these three subjects, we conducted a detailed review of existing documentation for each. We reviewed books, biographies, declassified government documents, court materials, as well as news articles and existing interviews with the whistleblowers. From these documents we constructed a detailed timeline of whistleblower actions and coded those behaviors as consistent or inconsistent with our theory.

We make two predictions about observable choices that whistleblowers make:

- They will go public with their accusations rather than use safe, internal channels;
- When they make their allegations, they will associate their name with them rather than leak information anonymously.

Through that detailed analysis, we confirmed that John Kiriakou and Daniel Ellsberg's cases well conform to our predictions. As we describe in the paper, while Drake initially deviated from our predictions, he ultimately conformed to them over time.

The purpose of the interviews was not to understand what the whistleblowers did *per se*, but rather why they did it. Looking through existing records, there was not enough information about Drake and Kiriakou's reasoning to either confirm or disconfirm our mechanism. In particular, there was no public account for why they opted to disclose through particular channels, how they weighed the decision to go public rather than private, and so forth. In contrast, Ellsberg has published multiple memoirs and has been outspoken about his experiences and the choices he made. Thus, we had a more complete picture about his reasoning from existing materials.

### **Interview Solicitation Rationale and Dealing with Non-compliance**

To address these gaps, we reached out to all three whistleblowers and requested to interview them about their experiences. John Kiriakou and Thomas Drake agreed to an interview. Daniel Ellsberg did not respond.

After repeated attempts to contact Dr. Ellsberg failed, we were faced with a choice: select a different subject or rely on existing sources. We chose to rely on existing sources for three reasons. First, Ellsberg may have a dispositional feature that led him to reject our interview request. That feature may also represent a different logic for whistleblowing. We did not want to only focus on those who responded, but rather the complete sample of whistleblowers. Second, there were a lot more public information about Ellsberg that we could rely on. Finally, no matter how well we craft interview questions, there is always the possibility of response bias. But past interviews and memoirs of Ellsberg were conducted by people who do not know our theory and were not trying to solicit responses that confirm it. Therefore, we can partially alleviate concerns about response bias in the other cases by showing that other sources report Ellsberg's reasoning was consistent with our mechanisms.

### **Interview Analysis Plan**

We constructed interview question templates to test our theory with each subject. The interview templates combined predictions from our theory, and the information we gathered during the review of existing documents. We included three kinds of questions: (1) questions to confirm our mechanism, (2) questions to confirm alternative mechanisms, and (3) questions to ensure subjects' answers were accurate and not subject to bias. Because we focus on Thomas Drake in the paper, the following discusses our analysis plan for that interview. The one for our interview with John Kiriakou was similar in most respects. Interested readers may contact the authors for additional details.

### **Questions Increasing Our Confidence Drake's Answers Were Accurate**

One major concern is that whistleblowers may not accurately remember their decision-making process. This is especially relevant in this case since Drake made his decisions over ten years ago. To address this concern, we searched for information Drake revealed in past interviews nearly a decade ago. We also searched for complex facts reported in government documents about Drake's case from that time period. We then tested Drake's recollection by asking him about these events. His responses in 2019 were remarkably consistent with previous responses he has given and government reports.

A second concern is response bias. It is possible that as we ask questions, Drake got a sense of the answers that we were looking for and started to tailor them accordingly. To address this, we asked Drake directly about alternative mechanisms to our theory. If Drake wanted to satiate the interviewer, he would have confirmed answers that worked against our theory. We provided Drake with several opportunities to do this. In no case did Drake affirm an alternative mechanism. As we discuss later on, these answers provide very strong support for our theory because they help us rule out alternatives.

A third concern is social desirability bias. Whistleblowers may present their behavior in ways that makes them sound normatively palatable. Political elites are especially subject

to these biases because there are usually many conflicting interpretations of their behavior. Elites typically face incentives to maintain a positive public image. In many cases, our theory suggests many positive attributes about whistleblowers. For example, we suggest that they can be altruistic, with high public service motivation, and they come forward to serve the greater good at high cost. When we ask Drake about these features, we want to be sure that they are true, and that Drake is not simply recounting the best version of events.

To address this concern, we asked Drake a number of questions that could prompt a socially desirable answer but one that would not be explicitly related to our theory. We reasoned that if Drake did not produce the socially desirable answer in these cases, that this was unlikely to be driving his answers in other areas.

First, we asked: “Do you think that you and other whistleblowers are unique in the national security community, or do you think most national security professionals carry similar value?” Drake’s response indicated that most national security professionals had high levels of public service motivation and want to do the right thing. He explicitly stated that he did not think he was unique.

Second, we asked Drake questions about the public benefit of his case. A socially desirable answer would claim Drake had a large impact on U.S. government surveillance programs. From reading about Drake’s case, we knew that his actions prompted an important discussion between Congress and the executive about surveillance and the limits of executive privilege. However, that conversation ultimately favored the executive. As a result, government surveillance did not end after Drake came forward. We suspected that Drake had hoped for more comprehensive reform than what was achieved. If Drake suffered from social desirability bias, he may have said he had a large impact, or accomplished what he wanted to accomplish. Yet, Drake gave a surprisingly honest answer. He said that his whistleblowing started an important discussion but did not end surveillance programs as he had hoped. As a result, he admitted that his actions did not have a large impact on reform (they had a large impact on credibly exposing government activities).

Third, we asked Drake generally if he thought about any private benefits he could accrue. Our theory suggests that whistleblowers that come forward typically do so for public benefits. However, they can also accrue some private value for coming forward. The most socially desirable case would be that Drake only wanted to stop executive abuse and that he did not consider private costs to be large or enjoy any benefits. Drake responded that he did realize that there were some private benefits to coming forward.

### **Questions Testing Our Mechanism**

Before we interviewed Drake and Kiriakou, we wrote out a detailed analysis plan for each. That plan included a link between our theoretical mechanism and the question we wrote out. It also included the answers that would confirm or disconfirm our theory. We report the following information from our analysis plan with Drake below:

- What we asked;
- Why we asked it;
- Answers that validate our theoretical mechanism;
- Answers that disconfirm our theoretical mechanism;
- Drake's answers and whether they (dis)confirm our theory.

The information we report is from the interview plan we wrote down before we conducted the interview. As a result, many of the questions are very similar, and all provide evidence for the same parts of our mechanism. We designed questions this way for three reasons. First, we were unsure what Drake would say, and therefore wrote out lines of questioning to account for the different answers he might give.

Second, we did not want to ask specific questions that would prime Drake to answer questions in a way that confirmed our theory. For example, if we asked, "Why did you choose to report to the Inspector General and not go straight to the press?" we are framing the

choice as a strategic problem with two outcomes. Instead, we ask the most general question, e.g. “Why did you choose to report abuse to the Inspector General?” We then see if Drake volunteers an answer that is consistent with the strategic logic of our model. Depending on his answer, we then move to more specific questions.

Third, we wanted to rule out plausible alternatives for Drake’s behavior. As a result, many of our questions explicitly ask Drake if his actions were motivated by a logic that would disconfirm our theory. Whenever we could think of a plausible alternative mechanism that explained Drake’s behavior, we asked him about it.

Q1 Before you decided to go public with your allegations, did you think about what might happen to you? If so, what were the costs you would face?

**Why we asked it.** We argue that whistleblowers choose between two reporting channels available to them. On the one hand, they can use internal reporting channels that will be safe (i.e. they will not face enormous costs) but ineffective (i.e. abuse is unlikely to stop). On the other, they can go public with their information. But if they do, they face enormous costs and risks. Alternatively, whistleblowers can do nothing (which is similar to ineffective internal reporting). Since our theory is rational, we predict that whistleblowers were aware of these different options and their costs and benefits of each before they came forward.

**Answers that support our theory.** If our theory is correct, subjects would have thought that they would risk jail time by going public and that they would lose their job. If our theory was incorrect, subjects might say that they did not know about the costs before they came forward or did not think the costs and risks were serious.

We focused only on one outcome (what happens when they come forward publicly) and only on the costs (rather than benefits) that they might face. We explicitly did not mention (1) benefits they might receive or (2) other options they might have. If subjects mentioned these things as part of their answers without any prompting, it would be strong evidence for

our theory because it suggests that subjects consider the costs and benefits of coming forward and the different options they have available in a manner consistent with our rational model. It also ensures that our inferences do not only follow from desirability bias.

In particular, it would be strong evidence in support of our theory if subjects mentioned that going public was the only way to having the abuse stop (i.e. benefits). It would also be strong evidence for our theory if subjects stated that their alternatives were that they could do nothing, or use internal channels and then lay out the costs and benefits for each.

**Answers that disconfirm our theory.** If subjects only answer the question directly, it is neither good nor bad for us. Follow on questions will ask about these things. However, if subjects listed alternatives other than doing nothing, or using internal channels, then it would suggest an alternative outcome we did not consider.

**What Drake said.** As reported in the main body, Drake explicitly stated that he understood the costs involved with coming forward. This supports our theory because it implies that Drake understood the costs and risks involved.

Drake further stated that he thought that internal channels would not be effective. This provides strong support for our theory because it implies that Drake instinctively framed his choice set as making an internal whistleblower complaint or external disclosure. Furthermore, his answer suggested that the core trade-off was between high personal cost that stopped abuse, and low personal cost that would not stop abuse. Drake provided no information that would disconfirm our theory.

Q2A When did you first raise concerns? [anticipating his answer would be in late 2001]

Q2B Who did you initially report your concerns to? [anticipating his answer to be the IG at NSA and DOD]

Q2C Why did you choose to initially report your concerns to [based on answer]?

**Why we asked it.** We argue that there is a specific kind of whistleblower, those who wish to see real policy change, that will come forward publicly with allegations of abuse. We expect that these whistleblowers use public channels because they want abuse to end and think that going public is the only way to end it. As described in the paper, Drake initially tried to make internal reports to the IG at the NSA and DOD. We wanted to understand why Drake deviated from our expectations. In particular, we thought if our theory was wrong, this would be the place to look.

We started with general questions because we did not want to prime Drake to think about his initial complaints to the IG as similar to his later reports. One possibility was that Drake thought of these all as part of the same reporting process and did not see a difference between his external and internal reporting. We wanted to make sure he clearly distinguished between these choices.

**Answers that support our theory.** Since we did not expect Drake to do this, we make no explicit prediction about why he did. However, we reasoned that it would be broadly consistent with our theory if Drake went internal because he did not know that the internal channels were ineffective. If he thought otherwise, he might try internal channels first hoping that they would work at low (or no) cost and only switch to external channels once he realized that internal channels would not be effective.

**Answers that would disconfirm our theory.** We thought of one reason that Drake would first report internally that would disconfirm our theory, namely that he thought doing so would immunize him from prosecution. If Drake thought that using internal channels before going external would provide him with total legal cover, it would disconfirm our theory. The reason is that our theory explicitly assumes that Drake's claims were credible because he exposed himself to costs.

**What Drake said.** As reported in the main paper, Drake’s answers explain his deviant behavior (relative to our expectations) in a way that is supportive of our theory. Drake reported that he initially went internal because he thought there was some chance that it would end abuse. It is largely consistent with our theory that a rational whistleblower who valued altruism would still try to avoid private costs.

Drake also provided a second answer that we did not anticipate. He stated that the rules were in place for good reason and that he wanted to follow them. This is consistent with our theory because it implies Drake understands the value of classification and does not want to release information for the sake of it. It also suggests that Drake did not have any underlying desire to shame or punish his employer. Drake provided no information that would disconfirm our theory.

Q3A When was the first time you provided information to someone outside of the US Government about events that were going on? [assuming Drake answers Siobhan Gorman at the Baltimore Sun 2006]

Q3B Why did you choose to go to [name reported in 3A]?

**Why we asked it.** Our theory predicts that whistleblowers go public despite the enormous costs of doing so in order to effect change. Given that Drake had previously tried internal channels and found them to be ineffective, we wanted to assess whether his rationale for going to the press conformed to our expectations.

**Answers that support our theory.** If our theory is correct, Drake would state that his goal was to end abuse and that he came to realize that this was the only way to effect that change. It would be stronger support for our theory if Drake explicitly stated that using internal channels were unlikely to go anywhere, or that staying quiet was not an option because abuse would never end.

**Answers that disconfirm our theory.** If our theory is incorrect, Drake might have said that he did not anticipate facing costs for coming forward. He might also have said that he believed the public had a right to know about what was happening.

**What Drake said.** As reported in the paper, Drake's answer was identical to the answer that supports our theory. Drake provided no information that would disconfirm our theory.

Q4A As you point out, you reported on surveillance programs to Inspectors General. You also point out the IG came back with a finding of waste, fraud, and abuse. If the NSA and the White House took the IG's report and then ended the programs, would you still have gone public with the information?

Q4B Do you think the public has an intrinsic right to know that the executive abused its power, as occurred in your case, if they correct it upon investigation?

**Why we asked it.** Our theory relies on the presumption that national security whistleblowers go to the press not because they believe the public has an inherent right to know state secrets but rather because they believe it is the only way to effect change. This question, then, gets at whether Drake would have sought out the press regardless of whether the policy he disliked was stopped (e.g. to alert the public).

**Answer that confirms our theory.** It would be strong support for our theory if Drake stated that there would be no need to go public with information once the executive changes their policy. Such an answer would confirm that Drake did not think the public had an intrinsic right to know the information. It would also be strong support for us if Drake explicitly stated that the government needs to keep secrets and the public should not have access to much national security information.

It would be reasonable support for us if Drake said going public would still be important because it would lead to the executive being punished and this would deter them from trying

things like this in the future. As a result, making the allegation public would prevent future abuse.

**Answer that disconfirms our theory.** If Drake said that the public has a right to know whenever the executive abuses secrecy, or that the public should punish the executive for any reason other than deterrence (such as revenge) it would suggest that going public served a different function than ending abuse. In this case, this would invalidate our theory (or at least suggest a different mechanism was important as well).

**What Drake said.** Drake stated that if the government ended the programs that concerned him, he never would have gone public with the information. Drake acknowledged that the government should be able to keep secrets and there are many pieces of information that are important to keep classified. Drake did not think that the public had an intrinsic right to know everything that the U.S. government was doing. Drake provided no information that would disconfirm our theory.

Q5A Initially when you released information to Siobhan Gorman, did you indicate your identity? [Anticipating he would say no.]

Q5B Why did you try to keep this information hidden from her?

Q5C Did you think that you could provide information to Ms. Gorman and keep your identity a secret?

**Why we asked it.** We argued that whistleblowers choose to go public and face enormous costs because going public lends credibility to their claims. From our pre-interview data collection, we noticed that Drake presented a tough test of our theory because he did not immediately link his name to his allegations. Instead, he provided information anonymously for a while, then provided information about his name.

We asked these questions as a tough test of our theory. We could not think of any reason that Drake would conceal his identity that would be consistent with our theory. By asking this question, we gave Drake the opportunity to invalidate our logic about why whistleblowers come forward.

**Answer that supports our theory.** We could think of no answer that supported our theory.

**Answer that disconfirms our theory.** Drake might have believed he could release information publicly without ever letting anyone know who he was. He may have thought getting the information out there anonymously could drive Congress to investigate and end abuse. If he never came forward, he would not suffer any costs. This would partially disconfirm our theory because it implies that Drake could not properly anticipate the consequences of his actions. That is, Drake was not aware that he would incur costs for coming forward. Further, it implies that costs are unnecessary to make his story credible.

**What Drake said.** Drake stated that he initially contacted Siobhan Gorman in secret using some anonymous email servers. Using those email servers, he pointed Gorman to unclassified information that, taken together, made clear the extent of the government surveillance programs. However, Drake knew that the government could quickly and easily figure out that it was him. Thus, he never thought that he was concealing his identity to avoid prosecution. Drake pointed to the very programs he was reporting on as evidence that he was under no illusions that he could keep his identity secret.

Drake said that he wanted to remain anonymous to see how Ms. Gorman would deal with the information. Once he realized that he could trust her to report accurately, he went to her office in person and revealed himself. He sat down with her and conducted several one-on-one interviews to provide context to the documents that he provided her. Consistent with our theory, Drake did not conceal his identity to avoid punishment.

Q6A Eventually, you revealed your identity to Ms. Gorman. Why did you choose to make your identity known?

**Why we asked it.** Our theory suggests that whistleblowers understand that it is very difficult to make their stories credible. We predict that they tie their name to their stories to lend credibility to the allegations. By making the allegations credible, abuse is more likely to end.

**Answer that supports our theory.** If our theory is correct, Drake would suggest that tying his name to the story lends considerable credibility. Thus, he came forward to increase the credibility of his claim. It would be especially strong support if Drake suggested that coming forward dramatically raised his own personal risk, and that those increased risks and costs are what drove the credibility of his claim.

**Answer that disconfirms our theory.** We could not think of a good reason that Drake would publicly come forward other than to lend credibility to his claims. We initially thought that Drake may have come forward as part of a strategy to protect himself from prosecution, but could not think of any way this would help him.

**What Drake said.** Consistent with our theory, Drake stated that he came forward to lend credibility to his allegations. He believed that having his name tied to the allegations made them intrinsically more credible. He also stated that through in-person interviews, he could clarify any questions that Gorman had.

In his answer, Drake stated that by coming forward he ran great risks, and Ms. Gorman understood that. But he did not explicitly link costs and risks to raised credibility. Thus, we asked him directly: do you think that by coming forward and accepting additional risk, your allegations were more credible and people were more likely to take notice. He responded, yes of course.

## Qualitative Appendix: Beyond Drake

In the main article, we explored the strategic implications of our mechanism as it pertains to whistleblowers by examining the case of Thomas Drake. Here, we illustrate our mechanism in two other critical cases. The first involves Daniel Ellsberg and his whistleblowing on the Vietnam War. The second involves John Kiriakou, a former Central Intelligence Agency officer who exposed the Bush administration’s use of torture in 2007. These two cases demonstrate that our argument extends to the pre-9/11 period (Ellsberg) and also to the period after Drake blew the whistle (Kiriakou). The latter involves an interview using the same techniques outlined in the article and the appendix.

### Ellsberg

Daniel Ellsberg began his career in the military in the mid-1950s. He later became an analyst at the RAND corporation before joining the government first as a special assistant to an assistant secretary of defense in the Office of the Secretary of Defense followed by a job with the State Department before finally returning to RAND. During his assignment at State, Ellsberg became distressed by the disparity between the government’s rosy assessment of the Vietnam War and the reality on the ground, some of which was written up in the highly-classified Pentagon Papers. Eventually, Ellsberg decided he could no longer stay silent on the matter (Stanger, 2019, 66-68).

Although nearly 50 years earlier, Ellsberg—similar to Drake—tried to handle the matter discreetly through quasi-internal channels, asking the “Chairman of the Senate Foreign Relations Committee, William Fulbright, to release select portions through hearings on the Senate floor under congressional privilege. This would have protected Ellsberg from prosecution” (Stanger, 2019, 69).<sup>27</sup> Fulbright declined Ellsberg’s request, as did other policymakers.

In March of 1971, Ellsberg approached the *New York Times*, who published the first batch

---

<sup>27</sup>It is worth noting that none of the significant whistleblower legislation, nor the House and Senate intelligence committees, existed during Ellsberg’s time.

of documents, along with his name, in June. After a government injunction not to publish any more, which the *Times* respected, Ellsberg passed the Pentagon Papers to a variety of other papers. In his recounting of the initial decision to go to the press, Ellsberg describes his frustration with alternative routes: “I became increasingly doubtful of getting anywhere with Congress.” He also notes that while he was skeptical the *New York Times* would take his story, he “was ready to take a bigger risk” (Ellsberg 2002, 375). Interestingly, he notes in a recent memoir that he declined to release additional top-secret documents about the nuclear program, which were also in his possession, because he feared that such information would overwhelm the news about Vietnam and stymie the prospect for change. As he apparently told one of his friends at the time, “Vietnam is where the bombs are falling right now... I think I have to give [Vietnam] as much of a run as I can first, for whatever difference it might make to shorten the war” (Ellsberg 2017).

While the newspapers survived a number of battles in court regarding their ability to publish the material, Ellsberg was hit with an Espionage Act charge. The government initially tried to discredit him by doing a psychological evaluation, but rejected it after it found that he acted as he did out of “a higher order of patriotism” and was “an ‘extremely intelligent and talented individual’ who ‘saw himself as having a special mission and bearing a special responsibility’ ” (Stanger, 2019, 71). Charges against Ellsberg were eventually dropped after revelations that the Nixon administration attempted to circumvent the FBI’s investigation by creating its own investigative unit.

### **Kiriakou**

John Kiriakou was “[t]he sixth individual the Obama administration prosecuted under the Espionage Act...” (Stanger, 2019, 109). The impetus for Kiriakou’s prosecution was the public revelation of the Bush administration’s enhanced interrogation program, which included waterboarding and other forms of torture, during a television interview on ABC. Kiriakou was eventually charged with violating the Intelligence Identities Protection Act and served nearly

three years in prison (Stanger, 2019, 109). As with Drake, the limited analyses of his case that do exist do not focus on the kinds of issues our model raises. Because of this, we decided to interview Kiriakou in May 2019. The details validate our assumptions and mechanisms.

To begin with, Kiriakou, like Drake, was motivated to join the CIA by a sense of patriotism and service. As he put it, “to hear my paternal grandfather tell this story, Franklin Roosevelt was personally waiting for him at Ellis Island to give him a job and a passport, and passage to Pittsburgh, Pennsylvania. He ... instilled in his own children this notion that public service was really the only logical thing to do. And so, when I was growing up, I only considered public service as a job... And I was very patriotic, and I wanted to see the world, I wanted to travel.”<sup>28</sup>

Kiriakou’s impetus for coming forward to blow the whistle on the torture program was, in his telling, driven by the same kinds of considerations that led him to join the CIA in the first place: “My first day (at the CIA) was January 8, 1990. And I raised my right hand in the air and swore to protect the Constitution against all enemies, foreign and domestic, and I meant it. It sickens me that I might have been the only person in the room that day that meant it. But I did, I meant it... And I believed deeply, I still do, that that program was criminal in nature. We have the Federal Torture Act, the United Nations Convention Against Torture... So, I knew I was taking a chance, but I was confident that I was on the right side.”

During the interview, we also asked Kiriakou about why he decided not to use internal channels and why he decided to go public rather than remain anonymous. His rationale for the former—including the fact that he saw what happened to Drake—is described in the article. His decision to forgo anonymity, however, was driven by his belief that exposing his identity was important to the credibility of the story and the prospect of achieving change: “I thought about remaining anonymous. And I thought, no, because that’s kind of the cowardly way to do it. People are going to deny it, criticize it. I wanted to put people on the spot. I wanted to make them put up if they were going to deny it... The public needs to know that

---

<sup>28</sup>Interview with authors, May 9, 2019.

it's me because I had access to the information and I knew what I was talking about... And I decided in the days leading up to that interview, that no matter what he asked me, I would tell the truth. And so, I did." Kiriakou also told us that he believed his willingness to expose himself to these risks increased his credibility.

## References

- Ellsberg, Daniel. 2017. *The Doomsday Machine: Confessions of a Nuclear War Planner*. New York: Bloomsbury Publishing.
- Ellsberg, Daniel. 2002. *Secrets: A Memoir of Vietnam and the Pentagon Papers*. New York: Penguin Books.
- Stanger, Allison. 2019. *Whistleblowers: Honesty in America From Washington to Trump*. New Haven and London: Yale University Press.